

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:31:02 ; Search time 35.5 Seconds
(without alignments)
35.769 Million cell updates/sec

Title: US-10-064-903-1
Perfect score: 29
Sequence: 1 HXXXHXXH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query		%		Description	
No.	Score	Match	Length	DB	ID		
1	23	79.3	30	22	AAO13629	Human polypeptide	
2	23	79.3	60	23	ABP10050	Human ORFX protein	
3	23	79.3	66	23	ABP33078	Human ORF2051 prot	
4	23	79.3	68	23	ABP04899	Human ORFX protein	
5	23	79.3	80	22	AAU60888	Propionibacterium	
6	23	79.3	92	22	AAO11260	Human polypeptide	
7	23	79.3	115	23	ABP03095	Human ORFX protein	
8	23	79.3	116	23	ABP01907	Human ORFX protein	
9	23	79.3	119	22	AAM25686	Human protein sequ	

10	23	79.3	132	22	AAB63312	Human breast cance
11	23	79.3	134	22	AAU67989	Propionibacterium
12	23	79.3	137	23	ABP02968	Human ORFX protein
13	23	79.3	141	22	AAO10851	Human polypeptide
14	23	79.3	167	22	ABB70944	Drosophila melanog
15	23	79.3	218	23	ABP31393	Human ORF366 prote
16	23	79.3	357	22	AAU64020	Propionibacterium
17	23	79.3	462	22	AAU42927	Propionibacterium
18	23	79.3	466	18	AAW09825	UDP-glucose:thiohy
19	23	79.3	508	22	ABB71345	Drosophila melanog
20	23	79.3	543	22	ABG22945	Novel human diagno
21	23	79.3	572	24	ABP77246	N. gonorrhoeae ami
22	23	79.3	635	18	AAW19920	Human Ksr' (kinase
23	23	79.3	880	22	ABB65766	Drosophila melanog
24	23	79.3	1078	24	ABP96069	Human protein kina
25	23	79.3	1133	22	ABB65544	Drosophila melanog
26	23	79.3	1187	22	ABB67666	Drosophila melanog
27	23	79.3	1518	24	ABJ18375	Breast specific re
28	23	79.3	1529	17	AAR97985	CORK potassium cha
29	23	79.3	1575	22	ABG27933	Novel human diagno
30	23	79.3	2424	22	ABB58924	Drosophila melanog
31	23	79.3	3502	22	ABB58382	Drosophila melanog
32	22	75.9	34	21	AAG07703	Arabidopsis thalia
33	22	75.9	37	22	AAM86640	Human immune/haema
34	22	75.9	41	21	AAB34597	Human secreted pro
35	22	75.9	49	22	ABB17194	Human nervous syst
36	22	75.9	50	23	ABP31516	Human ORF489 prote
37	22	75.9	52	23	ABG93190	S. cerevisiae BAX-
38	22	75.9	53	23	ABP07950	Human ORFX protein
39	22	75.9	55	22	AAU58208	Propionibacterium
40	22	75.9	56	23	ABP32940	Human ORF1913 prot
41	22	75.9	57	22	AAM78747	Human protein SEQ
42	22	75.9	59	22	AAM79731	Human protein SEQ
43	22	75.9	60	23	ABP05469	Human ORFX protein
44	22	75.9	61	22	AAO10714	Human polypeptide
45	22	75.9	61	22	AAG75865	Human colon cancer

ALIGNMENTS

RESULT 1

ID AAO13629 standard; Protein; 30 AA.

AC AAO13629;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 27521.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI93560.

XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX Claim 20; SEQ ID NO 27521; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 30 AA;
SQ
Query Match 79.3%; Score 23; DB 22; Length 30;
Best Local Similarity 37.5%; Pred. No. 4.3e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | | |
Db 18 HTHHTSH 25

RESULT 2
ABP10050
ID ABP10050 standard; Protein; 60 AA.

XX ABP10050;
AC
XX 25-JUN-2002 (first entry)
DT
XX Human ORFX protein sequence SEQ ID NO:20082.
DE
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
XX Homo sapiens.
OS
XX WO200192523-A2.
PN
XX 06-DEC-2001.
PD
XX 29-MAY-2001; 2001WO-US10836.
PF
XX 30-MAY-2000; 2000US-206132P.
PR
XX 29-AUG-2000; 2000US-228716P.
XX
XX (CURA-) CURAGEN CORP.
PA
XX Shinkets RA, Leach MD;
PI
XX WPI; 2002-106308/14.
DR
XX N-PSDB; ABN25802.
DR
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -

XX Disclosure; SEQ ID 20082; 1037pp; English.
PS
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, diabetes related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 60 AA;
SQ

Query Match 79.3%; Score 23; DB 23; Length 60;
Best Local Similarity 37.5%; Pred. No. 7.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | | |
Db 20 HHSSHTH 27

RESULT 3
ABP33078
ID ABP33078 standard; Protein; 66 AA.
XX
AC ABP33078;
XX
DT 09-JUL-2002 (first entry)
XX
DE Human ORF2051 protein, SEQ ID NO:4102.

XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide.

XX Homo sapiens.
OS
XX WO200190366-A2.
PN
XX 29-NOV-2001.
PD
XX 24-MAY-2001; 2001WO-US17076.
PF
XX 24-MAY-2000; 2000US-206690P.
PR

XX (CURA-) CURAGEN CORP.
XX Leach MD, Shimkets RA;
PI WPI; 2002-106200/14.
XX N-PSDB; ABN77104.
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation -
XX Claim 10; Page 1282-1283; 2508pp; English.
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX
SQ Sequence 66 AA;
Query Match 79.3%; Score 23; DB 23; Length 66;
Best Local Similarity 37.5%; Pred. No. 8e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 HXXXHXXH 8
Db 28 HHTTHTSH 35
RESULT 4
ABP04899
ID ABP04899 standard; Protein; 68 AA.
XX ABP04899;
AC
XX 25-JUN-2002 (first entry)
DT Human ORFX protein sequence SEQ ID NO:9780.
XX
DE Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW

KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX WO200192523-A2.
FN
XX
PD 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US10836.
PF
XX 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach MD;
PI WPI; 2002-106308/14.
XX N-PSDB; ABN20651.
DR Novel human polypeptides and polynucleotides useful for diagnosing,
DR preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders -
PS Disclosure; SEQ ID 9780; 1037pp; English.
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP0010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, immune deficiencies and disorders, infectious
CC storage disease, various immune deficiencies and disorders, rheumatoid
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 68 AA;
Query Match 79.3%; Score 23; DB 23; Length 68;
Best Local Similarity 37.5%; Pred. No. 8.2e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 HXXXHXXH 8
Db 9 HTHSHTAH 16
RESULT 5
AAU60888
ID AAU60888 standard; Protein; 80 AA.
XX
AC AAU60888;

XX DT 27-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein #21784.
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN WO2001181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US12865.
XX PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX DR WPI; 2001-616774/71.
DR N-PSDB; AAS59613.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX PS Example 1; SEQ ID NO 22083; 1069pp; English.
XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 80 AA;

Query Match 79.3%; Score 23; DB 22; Length 80;
Best Local Similarity 37.5%; Pred. No. 9.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 47 HSASHRTH 54

RESULT 6
AAO11260
ID AAO11260 standard; Protein; 92 AA.
XX
AC AAO11260;

XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 25152.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
PI WPI; 2001-514838/56.
XX DR N-PSDB; AAI91191.
XX PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX PS Claim 20; SEQ ID NO 25152; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 92 AA;

Query Match 79.3%; Score 23; DB 22; Length 92;
Best Local Similarity 37.5%; Pred. No. 1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 41 HTTAGHSH 48

RESULT 7
ABP03095
ID ABP03095 standard; Protein; 115 AA.
XX
AC ABP03095;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:6172.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN18847.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure; SEQ ID 6172; 1037pp; English.

CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP0010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 115 AA;

Query Match 79.3%; Score 23; DB 23; Length 115;
Best Local Similarity 37.5%; Pred. No. 1.3e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXXH 8
Db 45 HTTHTLH 52

RESULT 8 79.3%; Score 23; DB 23; Length 115;
ABP01907 37.5%; Pred. No. 1.3e+03;
ID ABP01907 standard; Protein; 116 AA.
XX
AC ABP01907;
XX

DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:3796.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN17659.

PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure; SEQ ID 3796; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP0010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 116 AA;

Query Match 79.3%; Score 23; DB 23; Length 116;
Best Local Similarity 37.5%; Pred. No. 1.3e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXXH 8
Db 46 HSTAHSBH 53

RESULT 8 79.3%; Score 23; DB 23; Length 116;
ABP01907 37.5%; Pred. No. 1.3e+03;
ID ABP01907 standard; Protein; 116 AA.
XX
AC ABP01907;
XX

XX AAU67989;
AC
XX 27-FEB-2002 (first entry)
DT
XX
DE Propionibacterium acnes immunogenic protein #28885.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-20841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59785.
XX
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris
XX
PS Example 1; SEQ ID No 29184; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 134 AA;

Query Match 79.3%; Score 23; DB 22; Length 134;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXXXXH 8
Db 40 HSTAHSBH 47

RESULT 12
ABP02968
ID ABP02968 standard; Protein; 137 AA.

XX
AC ABP02968;
XX
DT 25-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:5918.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach MD;
PI
PI N-PSDB; ABN18720.
DR
DR WPI; 2002-106308/14.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders
XX
PS Disclosure; SEQ ID 5918; 1037pp; English.
XX
XX The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
XX in the specification). ABN15762 to ABN27252 encode the human ORFX
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX storage disease, various immune deficiencies and disorders, infectious
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX bone degenerative disorders, or periodontal disease, and for gut
XX protection or regeneration and treatment of lung or liver fibrosis,
XX reperfusion injury in various tissues and conditions resulting from
XX systemic cytokine damage.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 137 AA;

Query Match 79.3%; Score 23; DB 23; Length 137;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXXXXH 8

KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide.
XX

OS Homo sapiens.

XX
PN WO200190366-A2.

XX
PD 29-NOV-2001.

XX
PF 24-MAY-2001; 2001WO-US17076.

XX
PR 24-MAY-2000; 2000US-206690P.

XX
PA (CURA-) CURAGEN CORP.

XX
PI Leach MD, Shimkets RA;

XX
DR WPI; 2002-106200/14.
DR N-PSDB; ABN75419.

XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation -
XX

PS Claim 10; Page 450; 2508pp; English.

XX

CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX

SQ Sequence 218 AA;

Query Match 79.3%; Score 23; DB 23; Length 218;
Best Local Similarity 37.5%; Pred. No. 2.1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8

Db 86 HTHSHAAH 93

Search completed: November 12, 2003, 09:35:06
Job time : 36.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:33:48 ; Search time 14 Seconds
(without alignments)
24.178 Million cell updates/sec

Title: US-10-064-903-1
Perfect score: 29
Sequence: 1 HXXXHXXH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	23	79.3	139	4	US-09-252-991A-32472
2	23	79.3	511	4	US-09-252-991A-22789
3	23	79.3	573	4	US-09-252-991A-24488
4	23	79.3	635	1	US-08-571-758-10
5	23	79.3	635	1	US-08-909-984A-10
6	23	79.3	635	1	US-08-909-983-10
7	22	75.9	151	2	US-08-858-767-30
8	22	75.9	151	2	US-08-863-028-30
9	22	75.9	179	4	US-09-615-192A-289
10	22	75.9	249	4	US-09-252-991A-22610
11	22	75.9	260	4	US-09-252-991A-20987
12	22	75.9	272	4	US-08-858-207A-447
13	22	75.9	323	4	US-09-328-352-6181
14	22	75.9	355	2	US-08-758-621-4
15	22	75.9	355	3	US-09-107-858-4
16	22	75.9	369	4	US-09-252-991A-25533
17	22	75.9	387	4	US-09-364-230-18
18	22	75.9	388	4	US-09-252-991A-31265
19	22	75.9	412	4	US-09-252-991A-26532
20	22	75.9	413	4	US-09-328-352-5589
21	22	75.9	431	1	US-08-311-023-2
22	22	75.9	447	4	US-09-252-991A-32122
23	22	75.9	481	4	US-09-252-991A-24508
24	22	75.9	504	4	US-09-252-991A-28224
25	22	75.9	516	4	US-09-252-991A-29719
26	22	75.9	533	4	US-09-252-991A-23560
27	22	75.9	542	4	US-09-107-532A-4858

28	22	75.9	559	2	US-08-756-317-7	Sequence 7, Appli
29	22	75.9	559	2	US-08-756-317-10	Sequence 10, Appl
30	22	75.9	559	4	US-09-672-749-2	Sequence 2, Appli
31	22	75.9	559	4	US-09-821-016-1	Sequence 1, Appli
32	22	75.9	582	4	US-09-252-991A-27626	Sequence 27626, A
33	22	75.9	582	4	US-09-252-991A-32678	Sequence 32678, A
34	22	75.9	637	4	US-09-252-991A-28952	Sequence 28952, A
35	22	75.9	706	4	US-09-193-562D-11	Sequence 25730, A
36	22	75.9	795	4	US-09-193-562D-12	Sequence 11, Appl
37	22	75.9	821	4	US-09-187-999-11	Sequence 11, Appl
38	22	75.9	834	4	US-08-844-057-2	Sequence 2, Appli
39	22	75.9	872	2	US-09-006-730-2	Sequence 2, Appli
40	22	75.9	872	4	US-08-785-071A-2	Sequence 2, Appli
41	22	75.9	876	1	US-09-012-872-2	Sequence 2, Appli
42	22	75.9	876	3	US-09-328-352-6626	Sequence 6626, Ap
43	22	75.9	893	4	US-09-134-001C-3600	Sequence 3600, Ap
44	22	75.9	897	4	US-09-193-562D-2	Sequence 2, Appli
45	22	75.9	905	4		

ALIGNMENTS

RESULT 1
US-09-252-991A-32472
; Sequence 32472, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32472
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32472

Query Match 79.3%; Score 23; DB 4; Length 139;
Best Local Similarity 37.5%; Pred. No. 5.7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 26 HTALHSSH 33

RESULT 2
US-09-252-991A-22789
; Sequence 22789, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22789
; LENGTH: 511
; TYPE: PRT

QY 1 HXXXHXXH 8
Db 17 HTSAHTQH 24

RESULT 6

US-08-909-983-10
; Sequence 10, Application US/08909983
; Patent No. 5747288
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Wasserman, David A.
; TITLE OF INVENTION: A No. 5747288el Protein Kinase Required for Ras
; TITLE OF INVENTION: Signal Transduction
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,983
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/571,758
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B96-010
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-909-983-10

Query Match 79.3%; Score 23; DB 1; Length 635;
Best Local Similarity 37.5%; Pred. No. 1.8e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 17 HTSAHTQH 24

RESULT 7

US-08-858-767-30
; Sequence 30, Application US/08858767
; Patent No. 5837468
; GENERAL INFORMATION:
; APPLICANT: WANG, Xun
; APPLICANT: DUVICK, Jonathan P.
; APPLICANT: BRIGGS, Steven P.
; TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING
; TITLE OF INVENTION: METHOD
; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,767
; FILING DATE: 19-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/481,687
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/325/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-858-767-30

Query Match 75.9%; Score 22; DB 2; Length 151;
Best Local Similarity 37.5%; Pred. No. 8.9e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 56 HAFHTDH 63

RESULT 8

US-08-863-028-30
; Sequence 30, Application US/08863028
; Patent No. 5853991
; GENERAL INFORMATION:
; APPLICANT: WANG, Xun
; APPLICANT: DUVICK, Jonathan P.
; APPLICANT: BRIGGS, Steven P.
; TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING
; TITLE OF INVENTION: METHOD
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,028
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,767

Query Match 79.3%; Score 23; DB 1; Length 635;
Best Local Similarity 37.5%; Pred. No. 1.8e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 17 HTSAHTQH 24

RESULT 7

US-08-858-767-30
; Sequence 30, Application US/08858767
; Patent No. 5837468
; GENERAL INFORMATION:
; APPLICANT: WANG, Xun
; APPLICANT: DUVICK, Jonathan P.
; APPLICANT: BRIGGS, Steven P.
; TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING
; TITLE OF INVENTION: METHOD
; NUMBER OF SEQUENCES: 39

```

; FILING DATE: 19-MAY-1997
; APPLICATION NUMBER: US 08/481,687
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/325/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-863-028-30

Query Match 75.9%; Score 22; DB 2; Length 151;
Best Local Similarity 37.5%; Pred. No. 8.9e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 56 HAFATDH 63

RESULT 9
US-09-615-192A-289
; Sequence 289, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 289
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-615-192A-289

Query Match 75.9%; Score 22; DB 4; Length 179;
Best Local Similarity 37.5%; Pred. No. 1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 170 HSIHSDH 177

RESULT 10
US-09-252-991A-22610
; Sequence 22610, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

```

```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22610
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22610

Query Match 75.9%; Score 22; DB 4; Length 249;
Best Local Similarity 37.5%; Pred. No. 1.3e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 60 HAAHHAH 67

RESULT 11
US-09-252-991A-20987
; Sequence 20987, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20987
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (78)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-20987

Query Match 75.9%; Score 22; DB 4; Length 260;
Best Local Similarity 37.5%; Pred. No. 1.3e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 113 HLAHRSH 120

RESULT 12
US-08-858-207A-447
; Sequence 447, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA

```


; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 447:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
; US-08-858-207A-447

Query Match 75.9%; Score 22; DB 4; Length 272;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 151 HTATHLLH 158

RESULT 13
US-09-328-352-6181
; Sequence 6181, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6181
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6181

Query Match 75.9%; Score 22; DB 4; Length 323;
Best Local Similarity 37.5%; Pred. No. 1.6e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 303 HIAQHASH 310

RESULT 14
US-08-758-621-4
; Sequence 4, Application US/08758621
; Patent No. 5846821
; GENERAL INFORMATION:
; APPLICANT: Guerinot, Mary Lou, and Eide, David J.

; TITLE OF INVENTION: Metal-Regulated Transporters and Uses Therefor
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,621
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/018,578
; FILING DATE: 29-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silveri, Jean M.
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: DCI-099CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-758-621-4

Query Match 75.9%; Score 22; DB 2; Length 355;
Best Local Similarity 37.5%; Pred. No. 1.7e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 178 H1H7HASH 185

RESULT 15
US-09-107-858-4
; Sequence 4, Application US/09107858
; Patent No. 6162900
; GENERAL INFORMATION:
; APPLICANT: Guerinot, Mary Lou et al.
; TITLE OF INVENTION: METAL-REGULATED TRANSPORTERS AND USES THEREFOR
; FILE REFERENCE: DCI-099CPDV
; CURRENT APPLICATION NUMBER: US/09/107,858
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 08/758,621
; EARLIER FILING DATE: 1996-11-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-107-858-4

Query Match 75.9%; Score 22; DB 3; Length 355;
Best Local Similarity 37.5%; Pred. No. 1.7e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 178 H1H7HASH 185

Wed Nov 12 16:14:07 2003

us-10-064-903-1.rai

Page 6

Search completed: November 12, 2003, 09:38:08
Job time : 15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: November 12, 2003, 09:35:43 ; Search time 23 Seconds
(without alignments)
59.739 Million cell updates/sec
Title: US-10-064-903-1
Perfect score: 29
Sequence: 1 HXXXHXXH 8
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 644079 seqs, 171749292 residues
Total number of hits satisfying chosen parameters: 644079
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match %	Description
1	23	79.3	890 15 US-10-156-761-14378
2	22	75.9	61 14 Sequence 14378, A
3	22	75.9	61 15 Sequence 228, Appl
4	22	75.9	69 9 US-10-106-698-6639
5	22	75.9	69 12 Sequence 6639, Ap
6	22	75.9	73 12 Sequence 35891, A
7	22	75.9	102 12 US-10-029-386-29728
8	22	75.9	104 10 US-10-238-075-517
9	22	75.9	110 9 US-10-231-417-489
10	22	75.9	110 9 Sequence 29728, A
11	22	75.9	114 9 Sequence 517, App
12	22	75.9	136 10 Sequence 489, App
13	22	75.9	161 15 US-09-864-761-37988
14	22	75.9	167 9 Sequence 1330, Ap
15	22	75.9	179 16 Sequence 35339, A
			Sequence 228, Appl
			Sequence 6639, Ap
			Sequence 35891, A
			Sequence 29728, A
			Sequence 517, App
			Sequence 489, App
			Sequence 1330, Ap
			Sequence 35339, A
			Sequence 46752, A
			Sequence 37988, A
			Sequence 74, Appl
			Sequence 4, Appli
			Sequence 34765, A
			Sequence 289, App

16	22	75.9	180	9	US-09-811-284-249	Sequence 249, Appl
17	22	75.9	193	9	US-09-191-687B-4	Sequence 4, Appli
18	22	75.9	193	15	US-10-228-796-4	Sequence 4, Appli
19	22	75.9	207	9	US-09-804-551B-42	Sequence 42, Appl
20	22	75.9	221	12	US-10-032-585-7060	Sequence 7060, Ap
21	22	75.9	340	9	US-09-971-361-10	Sequence 10, Appl
22	22	75.9	342	15	US-10-156-761-12399	Sequence 12399, A
23	22	75.9	352	15	US-10-232-563-2	Sequence 2, Appli
24	22	75.9	359	15	US-10-232-563-6	Sequence 6, Appli
25	22	75.9	359	15	US-10-232-563-7	Sequence 7, Appli
26	22	75.9	385	12	US-09-855-612-2	Sequence 2, Appli
27	22	75.9	385	14	US-10-139-262-2	Sequence 2, Appli
28	22	75.9	385	15	US-10-255-969-2	Sequence 2, Appli
29	22	75.9	397	10	US-09-925-300-1531	Sequence 1531, Ap
30	22	75.9	408	9	US-09-864-761-37954	Sequence 37954, A
31	22	75.9	420	12	US-10-160-764-84	Sequence 84, Appl
32	22	75.9	424	15	US-10-156-761-8087	Sequence 8087, Ap
33	22	75.9	483	10	US-09-905-999-20	Sequence 20, Appl
34	22	75.9	500	12	US-10-032-585-7530	Sequence 7530, Ap
35	22	75.9	527	15	US-10-128-714-3378	Sequence 3378, Ap
36	22	75.9	556	15	US-10-128-714-3561	Sequence 3561, Ap
37	22	75.9	556	15	US-10-128-714-8378	Sequence 8378, Ap
38	22	75.9	556	15	US-10-128-714-8561	Sequence 8561, Ap
39	22	75.9	559	9	US-09-821-016-1	Sequence 1, Appli
40	22	75.9	559	9	US-09-820-952A-1	Sequence 1, Appli
41	22	75.9	559	9	US-09-820-721A-1	Sequence 1, Appli
42	22	75.9	559	10	US-09-364-847-21	Sequence 21, Appl
43	22	75.9	559	15	US-10-218-519-1	Sequence 1, Appli
44	22	75.9	559	15	US-10-259-632-1	Sequence 1, Appli
45	22	75.9	559	15	US-10-266-787-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-156-761-14378
; Sequence 14378, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14378
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14378

Query Match 79.3%; Score 23; DB 15; Length 890;
Best Local Similarity 37.5%; Pred. No. 5.8e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
| | |
Db 573 HSATHLTH 580

RESULT 2
US-10-001-835-228

Sequence 228, Application US/10001835
Publication No. US20020160387A1
GENERAL INFORMATION:
APPLICANT: Salceda, Susana
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Cafferkey, Robert
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
FILE REFERENCE: DEX-0277
CURRENT APPLICATION NUMBER: US/10/001,835
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: 60/249,997
PRIOR FILING DATE: 2000-11-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: PatentIn version 3.1
SEQ ID NO 228
LENGTH: 61
TYPE: PRT
ORGANISM: Homo sapien
US-10-001-835-228

Query Match 75.9%; Score 22; DB 14; Length 61;
Best Local Similarity 37.5%; Pred. No. 1.3e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
DB 22 HRSTHQA 29

RESULT 3
US-10-106-698-6639
Sequence 6639, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 6639
LENGTH: 61
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (24)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (28)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (49)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (53)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (61)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6639

Query Match 75.9%; Score 22; DB 15; Length 61;

Best Local Similarity 37.5%; Pred. No. 1.3e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 HXXXHXXH 8
DB 41 HASDHFAH 48
RESULT 4
US-09-864-761-35891
Sequence 35891, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 35891
LENGTH: 69
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009743.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
OTHER INFORMATION: EST_HUMAN HIT: AA641863.1, EVALUE 7.10e-01
OTHER INFORMATION: SWISSPROT HIT: P04929, EVALUE 3.30e+00

US-09-864-761-35891

Query Match 75.9%; Score 22; DB 9; Length 69;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | | |
Db 51 HTVQHTSH 58

RESULT 5
US-10-029-386-29728
; Sequence 29728, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29728
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR11.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.67
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P22932, EVALUUE 1.20e+00
US-10-029-386-29728

Query Match 75.9%; Score 22; DB 12; Length 69;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | | |
Db 52 HDAHSGH 59

RESULT 6
US-10-238-075-517
; Sequence 517, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of thei
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 517
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-517

Query Match 75.9%; Score 22; DB 12; Length 73;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | | |

Db 15 HSHQHTAH 22
| | | | |

RESULT 7
US-10-231-417-489
; Sequence 489, Application US/10231417
; Publication No. US20030176681A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: PZ019P1
; CURRENT APPLICATION NUMBER: US/10/231,417
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/296,622
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 619
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 489
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-417-489

Query Match 75.9%; Score 22; DB 12; Length 102;
Best Local Similarity 37.5%; Pred. No. 1.8e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | | |
Db 56 HTHHTGH 63

RESULT 8
US-09-764-864-1330
; Sequence 1330, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1330
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (65)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (77)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (102)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1330

Query Match 75.9%; Score 22; DB 10; Length 104;
Best Local Similarity 37.5%; Pred. No. 1.9e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | | |
Db 81 HTRAHTAH 88

RESULT 9
US-09-864-761-35339

```
; Sequence 35339, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35339
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000507.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: Q30201, EVALUE 9.00e-16
; OTHER INFORMATION: EST_HUMAN HIT: BE877225.1, EVALUE 8.00e-54
US-09-864-761-35339
```

```
Query Match 75.9%; Score 22; DB 9; Length 110;
Best Local Similarity 37.5%; Pred. No. 1.9e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 HXXXHXXH 8
Db 90 HSGNHSTH 97

RESULT 10
US-09-864-761-46752
; Sequence 46752, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46752
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007389.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EST_HUMAN HIT: AA305279.1, EVALUE 4.00e-23
```

```

; OTHER INFORMATION: SWISSPROT HIT: Q26609, EVALUE 8.00e+00
US-09-864-761-46752

Query Match          75.9%; Score 22; DB 9; Length 110;
Best Local Similarity 37.5%; Pred. No. 1.9e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 HXXXHXXH 8
Db      55 HQATHSRH 62

RESULT 11
US-09-864-761-37988
; Sequence 37988, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37988
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO D84394.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.9
```

```

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.99
; OTHER INFORMATION: EST_HUMAN HIT: BE743982.1, EVALUE 8.00e-54
; OTHER INFORMATION: SWISSPROT HIT: Q30201, EVALUE 9.00e-16
US-09-864-761-37988

Query Match          75.9%; Score 22; DB 9; Length 114;
Best Local Similarity 37.5%; Pred. No. 2e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 HXXXHXXH 8
Db      94 HSGNHSTH 101

RESULT 12
US-09-893-737-74
; Sequence 74, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-74

Query Match          75.9%; Score 22; DB 10; Length 136;
Best Local Similarity 37.5%; Pred. No. 2.2e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 HXXXHXXH 8
Db      109 HGTAAHARH 116

RESULT 13
US-10-213-880-4
; Sequence 4, Application US/10213880
; Publication No. US20030088083A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Rasco-Gaunt, Sonriza
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Metal-Binding Proteins
; FILE REFERENCE: BB1513 US NA
; CURRENT APPLICATION NUMBER: US/10/213,880
; CURRENT FILING DATE: 2002-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/310,522
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Momordica charantia
US-10-213-880-4

Query Match          75.9%; Score 22; DB 15; Length 161;
```



```
Best Local Similarity 37.5%; Pred. No. 2.5e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 15 HSHSHSGH 22

RESULT 14
US-09-864-761-34765
; Sequence 34765, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34765
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006371.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
```

```
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EST_HUMAN HIT: BE958003.1, EVALUE 4.60e+00
US-09-864-761-34765

Query Match 75.9%; Score 22; DB 9; Length 167;
Best Local Similarity 37.5%; Pred. No. 2.6e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 112 HMHTHTSH 119

RESULT 15
US-10-174-693-289
; Sequence 289, Application US/10174693
; Publication No. US20030131373A1
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003C5
; CURRENT APPLICATION NUMBER: US/10/174,693
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 289
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-10-174-693-289

Query Match 75.9%; Score 22; DB 16; Length 179;
Best Local Similarity 37.5%; Pred. No. 2.7e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 170 HSIHSDH 177

Search completed: November 12, 2003, 09:44:49
Job time : 23 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 12, 2003, 09:33:02 ; Search time 13.5 Seconds
(without alignments)
56.989 Million cell updates/sec

Title: US-10-064-903-1
Perfect score: 29
Sequence: 1 HXXXHXXH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: Pirl: *
2: Pirl2: *
3: Pirl3: *
4: Pirl4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	23	79.3	152	2	C72662	hypothetical prote
2	23	79.3	177	2	T26468	hypothetical prote
3	23	79.3	240	2	F82790	GMP synthase XF056
4	23	79.3	327	2	AC2120	cytochrome c oxida
5	23	79.3	342	2	T15850	hypothetical prote
6	23	79.3	382	2	T35709	hypothetical prote
7	23	79.3	472	2	T27755	hypothetical prote
8	23	79.3	508	2	S59870	fork head domain p
9	23	79.3	510	2	S55124	probable membrane
10	23	79.3	606	2	B69805	conserved hypothet
11	23	79.3	826	2	T46060	hypothetical prote
12	23	79.3	826	2	T46061	hypothetical prote
13	22	75.9	52	2	S63324	hypothetical prote
14	22	75.9	61	2	AC0287	hypothetical prote
15	22	75.9	121	2	D82711	hypothetical prote
16	22	75.9	135	2	I49275	hypothetical prote
17	22	75.9	144	2	H75636	protein kinase STY
18	22	75.9	177	2	S65780	transposase-relate
19	22	75.9	198	2	B83717	glycine/proline-ri
20	22	75.9	208	2	T35454	NADP-dependent alc
21	22	75.9	237	2	S19103	hypothetical prote
22	22	75.9	263	2	G75590	hypothetical prote
23	22	75.9	306	2	I49068	protein kinase STY
24	22	75.9	312	2	T27004	hypothetical prote
25	22	75.9	316	2	D71375	probable ABC trans
26	22	75.9	325	2	T44782	exbB protein [limp
27	22	75.9	337	1	A42654	alcohol dehydrogen
28	22	75.9	339	1	S45605	alcohol dehydrogen
29	22	75.9	339	1	S47643	alcohol dehydrogen

30	22	75.9	340	2	T37030	alcohol dehydrogen
31	22	75.9	341	2	E83340	hypothetical prote
32	22	75.9	345	2	T16935	hypothetical prote
33	22	75.9	355	2	T52183	zinc transporter Z
34	22	75.9	364	2	JC5800	peptidylglycine mo
35	22	75.9	383	2	A55739	(MIC) protein MHC
36	22	75.9	384	2	T23604	hypothetical prote
37	22	75.9	394	2	E87606	hypothetical prote
38	22	75.9	403	2	C96757	hypothetical prote
39	22	75.9	416	2	A32947	filaggrin precursor
40	22	75.9	419	2	JQ2254	farnesyl-diphospha
41	22	75.9	420	2	G95107	gamma-glutamyl pho
42	22	75.9	420	2	A97976	glutamate-5-semial
43	22	75.9	424	2	T01383	GTPase-activating
44	22	75.9	427	2	I51580	XFKH2 protein - Af
45	22	75.9	440	2	B71293	hypothetical prote

ALIGNMENTS

RESULT 1
C72662
hypothetical protein APE0723 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: C72662
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takal
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon., Aeropy.
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: C72662
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-152 <KAW>
A;Cross-references: DDBJ:AP000060; NID:G5104188; PIDN:BAA79699.1; PID:d1043485; PID:G51.
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0723

Query Match 79.3%; Score 23; DB 2; Length 152;
Best Local Similarity 37.5%; Pred. No. 3e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 15 HSTIHAH 22

RESULT 2
T26468
hypothetical protein Y11D7A.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26468
R;Steward, C.
Submitted to the EMBL Data Library, October 1998
A;Reference number: Z20218
A;Accession: T26468
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-177 <WIL>
A;Cross-references: EMBL:AL032632; PIDN:CAA21589.1; GSPDB:GN00022; CESP:Y11D7A.1
A;Experimental source: clone Y11D7A
C;Genetics:
A;Gene: CESP:Y11D7A.1
A;Map position: 4
A;Introns: 48/1; 102/1; 128/1

Query Match 79.3%; Score 23; DB 2; Length 177;
Best Local Similarity 37.5%; Pred. No. 3.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 140 HTRVHSSH 147

RESULT 3
F82790
GMP synthase XF0560 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: F82790
R;anonymouse, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82790
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-240 <SIM>
A;Cross-references: GB:AE003903; GB:AE003849; NID:g9105416; PIDN:AAF83370.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorrry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0560

Query Match 79.3%; Score 23; DB 2; Length 240;
Best Local Similarity 37.5%; Pred. No. 4.5e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 142 HFSAHATH 149

RESULT 4
AC2120
cytochrome c oxidase chain II [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AC2120
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC2120
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-327 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAE74213.1; PID:gi7131606; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: coxB
C;Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C;Keywords: copper; electron transfer; membrane-associated complex; respiratory chain
F;214,249,253,260/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F;249,251,253,257/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted

F;251/Binding site: magnesium (Glu) (shared with chain I) #status predicted
Query Match 79.3%; Score 23; DB 2; Length 327;
Best Local Similarity 37.5%; Pred. No. 6e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 126 HASAHVAH 133

RESULT 5
T15850
hypothetical protein C56C10.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15850
R;Fulton, L.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of C. elegans cosmid C56C10.
A;Reference number: Z18417
A;Accession: T15850
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-342 <FUL>
A;Cross-references: EMBL:U29488; NID:g868238; PID:g868248; PIDN:AAA68778.1; CESP:C56C10
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C56C10.10
A;Introns: 51/2; 144/2; 204/3; 241/3; 295/3

Query Match 79.3%; Score 23; DB 2; Length 342;
Best Local Similarity 37.5%; Pred. No. 6.2e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 128 HSHAHTTH 135

RESULT 6
T35709
hypothetical protein SC7H1.14 SC7H1.14 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35709
R;Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z21548
A;Accession: T35709
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-382 <MUR>
A;Cross-references: EMBL:AL021411; PIDN:CAA16201.1; GSPDB:GN00070; SCOEDB:SC7H1.14
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC7H1.14

Query Match 79.3%; Score 23; DB 2; Length 382;
Best Local Similarity 37.5%; Pred. No. 6.9e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 370 HAARHAAH 377

RESULT 7
T27755
hypothetical protein ZK1320.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Dec-2002
C;Accession: T27755

R;Berks, M.
Submitted to the EMBL Data Library, December 1994
A;Reference number: Z20414
A;Accession: T27755
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-472 <WIL>
A;Cross-references: EMBL:Z46934; PIDN:CAA87047.1; GSPDB:GN00020; CESP:ZK1320.9
A;Experimental source: clone ZK1320
C;Genetics:
A;Gene: CESP:ZK1320.9
A;Map position: 2
A;Introns: 19/2; 55/1; 106/1; 160/1; 186/1; 323/1; 411/3
C;Superfamily: acetyl-CoA hydrolase

Query Match 79.3%; Score 23; DB 2; Length 472;
Best Local Similarity 37.5%; Pred. No. 8.3e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 198 HTTVHSSH 205

RESULT 8
S59870
fork head domain protein crocodile - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C;Accession: S59870; A46178
R;Haecker, U.; Kaufmann, E.; Hartmann, C.; Juergens, G.; Knoechel, W.; Jaeckle, H.
EMBO J. 14, 5306-5317, 1995
A;Title: The Drosophila fork head domain protein crocodile is required for the establish
A;Reference number: S59870; MUID:96080166; PMID:7489720
A;Accession: S59870
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-508 <HAE>
R;Hacker, U.; Grossniklaus, U.; Gehring, W.J.; Jackle, H.
Proc. Natl. Acad. Sci. U.S.A. 89, 8754-8758, 1992
A;Title: Developmentally regulated Drosophila gene family encoding the fork head domain.
A;Reference number: A46178; MUID:92409595; PMID:1356269
A;Accession: A46178
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 55-182 <HAC>
A;Cross-references: GB:M96440; NID:g157425; PIDN:AAF02177.1; PID:g6042185
A;Note: sequence extracted from NCBI backbone (NCBIP:114222)
C;Genetics:
A;Gene: croc
A;Cross-references: FlyBase:FBgn0014143
C;Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F;70-161/Domain: fork head DNA-binding domain homology <FHD>

Query Match 79.3%; Score 23; DB 2; Length 508;
Best Local Similarity 37.5%; Pred. No. 8.9e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 207 HMAAHAAH 214

RESULT 9
S55124
probable membrane protein YMR177w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YMR177w (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 06-Feb-1998
C;Accession: S55124
R;Churcher, C.M.
submitted to the EMBL Data Library, June 1995
A;Reference number: S55118

A;Accession: S55124
A;Molecule type: DNA
A;Residues: 1-510 <CHU>
A;Cross-references: EMBL:Z49808; NID:g854440; PID:g854447; MIPS:YMR177w
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:MMT1
A;Cross-references: SGD:S0004789; MIPS:YMR177w
A;Map position: 13R
C;Keywords: transmembrane protein
F;168-184/Domain: transmembrane #status predicted <TM1>
F;232-248/Domain: transmembrane #status predicted <TM2>
F;332-348/Domain: transmembrane #status predicted <TM3>
F;350-366/Domain: transmembrane #status predicted <TM4>

Query Match 79.3%; Score 23; DB 2; Length 510;
Best Local Similarity 37.5%; Pred. No. 8.9e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 137 HTHSHAAH 144

RESULT 10
B69805
conserved hypothetical protein yfix - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: B69805
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bette
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69805
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-606 <KUN>
A;Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CAB12672.1; PID:ell82833
A;Experimental source: strain 168
C;Genetics:
A;Gene: yfix

Query Match 79.3%; Score 23; DB 2; Length 606;
Best Local Similarity 37.5%; Pred. No. 1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 244 HSTSHITH 251

RESULT 11
T46060
hypothetical protein T18N14.20 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46060
R;Delseny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.;
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23013

A;Accession: T46060
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-826
A;Cross-references: EMBL:AL132968
A;Experimental source: cultivar Columbia; BAC clone T18N14
C;Genetics:
A;Map position: 3
A;Introns: 476/3; 796/2
A;Note: T18N14.20

Query Match 79.3%; Score 23; DB 2; Length 826;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 446 HTYAHSSH 453

RESULT 12
T46061
hypothetical protein T18N14.30 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46061
R;Delseny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.;
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23013
A;Accession: T46061
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-826
A;Cross-references: EMBL:AL132968
A;Experimental source: cultivar Columbia; BAC clone T18N14
C;Genetics:
A;Map position: 3
A;Introns: 476/3; 796/2
A;Note: T18N14.30

Query Match 79.3%; Score 23; DB 2; Length 826;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 446 HTYAHSSH 453

RESULT 13
S63324
hypothetical protein YNL338w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein N0170
C;Species: Saccharomyces cerevisiae
C;Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
C;Accession: S63324
R;Obermaier, B.; Piravandi, E.; Rinke, M.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63317
A;Accession: S63324
A;Molecule type: DNA
A;Residues: 1-52 <OBE>
A;Cross-references: EMBL:Z71614; NID:gl302466; PID:e239576; PID:gl302467; GSPDB:GN00014;
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YNL338w
A;Cross-references: SGD:S0005282
A;Map position: 14L

Query Match 75.9%; Score 22; DB 2; Length 52;
Best Local Similarity 37.5%; Pred. No. 1.7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 38 HTHHTHH 45

RESULT 14
AC0287
hypothetical protein YPO2354 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AC0287
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.E.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0287
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-61 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC91159.1; PID:gl5980351; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2354

Query Match 75.9%; Score 22; DB 2; Length 61;
Best Local Similarity 37.5%; Pred. No. 2e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 47 HHTHTSH 54

RESULT 15
D82711
hypothetical protein XF1205 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: D82711
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: D82711
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-121 <SIM>
A;Cross-references: GB:AE003954; GB:AE003849; NID:g9106165; PIDN:AAF84015.1; GSPDB:GN0
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer,
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fro
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Lai
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawas
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silve
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1205

Query Match 75.9%; Score 22; DB 2; Length 121;
Best Local Similarity 37.5%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8

Db | | |
 63 HTFTHTEH 70

Search completed: November 12, 2003, 09:37:28
Job time : 15.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search, using sw model
Run on: November 12, 2003, 09:31:57 ; Search time 10 Seconds
(without alignments)
37.621 Million cell updates/sec

Title: US-10-064-903-1
Perfect score: 29
Sequence: 1 HXXXHXXH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	23	79.3	508	1	CROC DROME	P32027 drosophila
2	23	79.3	510	1	YM43_YEAST	Q03218 saccharomyc
3	23	79.3	880	1	BRC4_DROME	Q24206 drosophila
4	23	79.3	890	1	SYA_STRCP	Q9kxp9 streptomyc
5	23	79.3	1509	1	GSRI_HUMAN	Q9nzm4 homo sapien
6	22	75.9	52	1	YN78_YEAST	P53820 saccharomyc
7	22	75.9	316	1	Y034_TREPA	O83077 treponema p
8	22	75.9	337	1	ADH1_BACST	P12311 bacillus st
9	22	75.9	339	1	ADH2_BACST	P42327 bacillus st
10	22	75.9	339	1	ADH3_BACST	P42328 bacillus st
11	22	75.9	416	1	FILA_HUMAN	P20930 homo sapien
12	22	75.9	419	1	PFTB_PEA	Q04903 pismo sativ
13	22	75.9	420	1	PROA_NEIMB	Q9jzgj neisseria m
14	22	75.9	420	1	PROA_STRPN	Q97r94 streptococc
15	22	75.9	427	1	FKH2_XENLA	P32315 xenopus lae
16	22	75.9	440	1	Y693_TREPA	O83691 treponema p
17	22	75.9	472	1	SK14_DROME	P40656 drosophila
18	22	75.9	483	1	CLK1_MOUSE	P22518 mus musculu
19	22	75.9	539	1	DOP2_DROME	Q24563 drosophila
20	22	75.9	559	1	PHAA_PSEOL	P26494 pseudomonas
21	22	75.9	590	1	SYT3_HUMAN	Q9bqg1 homo sapien
22	22	75.9	594	1	SYA_BORBU	O51238 borrelia bu
23	22	75.9	596	1	FRDA_SHEFR	Q02469 shewanella
24	22	75.9	605	1	SYA_TREPA	O83980 treponema p
25	22	75.9	679	1	TKT1_YEAST	P23254 saccharomyc
26	22	75.9	787	1	AGL2_BACTQ	Q9f234 bacillus th
27	22	75.9	842	1	SYA_CAMJE	Q9pi05 campylobact
28	22	75.9	860	1	SYA_VIBCH	Q56648 vibrio chol
29	22	75.9	860	1	SYA_VIBPA	Q871r3 vibrio para
30	22	75.9	860	1	SYA_VIBVU	Q8dc49 vibrio vuln
31	22	75.9	863	1	SYA_THEMEA	Q9x1b6 thernotoga
32	22	75.9	867	1	SYA_AQUAE	O67323 aquifex aeo
33	22	75.9	867	1	SYA_FUSNN	Q8rfj8 fusobacteri

34	22	75.9	871	1	SYA_AQUPY	Q9xdm3 aquifex pyr
35	22	75.9	872	1	SYA_LACLA	Q9cew0 lactococcus
36	22	75.9	872	1	SYA_STRP8	Q8p0e6 streptococc
37	22	75.9	872	1	SYA_STRPN	Q97q48 streptococc
38	22	75.9	872	1	SYA_STRPY	Q99z57 streptococc
39	22	75.9	873	1	SYA_WIGBR	Q8d2w8 wiggleswort
40	22	75.9	874	1	SYA_HAEIN	P43815 haemophilus
41	22	75.9	874	1	SYA_NEIMA	Q9jtg4 neisseria m
42	22	75.9	874	1	SYA_NEIMB	Q9jyg6 neisseria m
43	22	75.9	874	1	SYA_PASMU	P57933 pasteurella
44	22	75.9	874	1	SYA_PSEAE	Q9i553 pseudomonas
45	22	75.9	875	1	SYA_YERPE	Q8zbt8 yersinia pe

ALIGNMENTS

RESULT 1
CROC_DROME
ID CROC_DROME STANDARD; PRT; 508 AA.
AC P32027; Q9VP32;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fork head domain protein crocodile (FKH protein FD1).
GN CROC OR FD78E OR FD1 OR CG5069.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=96080166; PubMed=7489720;
RA Haecker U., Kaufmann E., Hartmann C., Juergens G., Knoechel W.,
RA Jaeckle H.;
RT "The Drosophila fork head domain protein crocodile is required for
the establishment of head structures."
RL EMBO J. 14:5306-5317(1995).
[2]
SEQUENCE FROM N.A.
STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE OF 55-182 FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND
RP DEVELOPMENTAL STAGE.
RX MEDLINE=92409595; PubMed=1356269;
RA Haecker U., Grossniklaus U., Gehring W.J., Jaeckle H.;
RT "Developmentally regulated Drosophila gene family encoding the fork
RT head domain.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992).
RN
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF HEAD STRUCTURES.
CC REQUIRED TO FUNCTION AS AN EARLY PATTERNING GENE IN THE ANTERIOR-
CC MOST BLASTODERM HEAD SEGMENT ANLAGE AND FOR THE ESTABLISHMENT OF A
CC SPECIFIC HEAD SKELETAL STRUCTURE THAT DERIVES FROM THE NON-
CC ADJACENT INTERCALARY SEGMENT AT A LATER STAGE OF EMBRYOGENESIS.
CC BINDS THE CONSENSUS DNA SEQUENCE 5'-(AG)TAAA[TC]A-3'.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY BLASTODERM EMBRYOS IN
CC ANTERIOR AND POSTERIOR GUT PRECURSORS, AND, LATER IN A SUBSET OF
CC CELLS IN CENTRAL NERVOUS SYSTEM.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYOGENESIS, MAXIMALLY
CC DURING THE 5-12 HOUR PERIOD.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; S80254; AAB35643.1; -.
DR EMBL; AE003594; AAF51727.1; -.
DR EMBL; M96440; AAF02177.1; -.
DR PIR; S59870; S59870.
DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T02291; -.
DR FlyBase; FBgn0014143; croc.
DR GO; GO:0005634; C:nucleus; IDA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 34 40 POLY-ALA.
FT DNA_BIND 69 160 FORK-HEAD.
FT DOMAIN 161 165 POLY-ARG.
FT DOMAIN 301 304 POLY-ALA.
FT DOMAIN 377 380 POLY-ASN.
FT DOMAIN 389 403 POLY-GLY.
FT DOMAIN 452 461 POLY-ALA.
FT DOMAIN 466 473 POLY-HIS.
FT VARIANT 122 122 L -> F (IN ALLELE CROC-75-3).
FT VARIANT 453 453 A -> V (IN ALLELE CROC-75-3).
SQ SEQUENCE 508 AA; 54516 MW; 2EPED1D8F63016D6 CRC64;

Query Match 79.3%; Score 23; DB 1; Length 508;
Best Local Similarity 37.5%; Pred. No. 4.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE OF 55-182 FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND
RP DEVELOPMENTAL STAGE.
RX MEDLINE=92409595; PubMed=1356269;
RA Haecker U., Grossniklaus U., Gehring W.J., Jaeckle H.;
RT "Developmentally regulated Drosophila gene family encoding the fork
RT head domain.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992).
RN
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF HEAD STRUCTURES.
CC REQUIRED TO FUNCTION AS AN EARLY PATTERNING GENE IN THE ANTERIOR-
CC MOST BLASTODERM HEAD SEGMENT ANLAGE AND FOR THE ESTABLISHMENT OF A
CC SPECIFIC HEAD SKELETAL STRUCTURE THAT DERIVES FROM THE NON-
CC ADJACENT INTERCALARY SEGMENT AT A LATER STAGE OF EMBRYOGENESIS.
CC BINDS THE CONSENSUS DNA SEQUENCE 5'-(AG)TAAA[TC]A-3'.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY BLASTODERM EMBRYOS IN
CC ANTERIOR AND POSTERIOR GUT PRECURSORS, AND, LATER IN A SUBSET OF
CC CELLS IN CENTRAL NERVOUS SYSTEM.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYOGENESIS, MAXIMALLY
CC DURING THE 5-12 HOUR PERIOD.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; S80254; AAB35643.1; -.
DR EMBL; AE003594; AAF51727.1; -.
DR EMBL; M96440; AAF02177.1; -.
DR PIR; S59870; S59870.
DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T02291; -.
DR FlyBase; FBgn0014143; croc.
DR GO; GO:0005634; C:nucleus; IDA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 34 40 POLY-ALA.
FT DNA_BIND 69 160 FORK-HEAD.
FT DOMAIN 161 165 POLY-ARG.
FT DOMAIN 301 304 POLY-ALA.
FT DOMAIN 377 380 POLY-ASN.
FT DOMAIN 389 403 POLY-GLY.
FT DOMAIN 452 461 POLY-ALA.
FT DOMAIN 466 473 POLY-HIS.
FT VARIANT 122 122 L -> F (IN ALLELE CROC-75-3).
FT VARIANT 453 453 A -> V (IN ALLELE CROC-75-3).
SQ SEQUENCE 508 AA; 54516 MW; 2EPED1D8F63016D6 CRC64;

Query Match 79.3%; Score 23; DB 1; Length 508;
Best Local Similarity 37.5%; Pred. No. 4.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 207 HMAAHAAH 214

RESULT 2
YM43_YEAST STANDARD; PRT; 510 AA.
ID YM43_YEAST
AC Q03218;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 56.2 kDa protein in SIP18-SPT21 intergenic region.
GN YMR177W OR YM8010.07.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";
RL Nature 387:90-93(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: STRONG, TO YEAST YPL224C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z49808; CAA89910.1; -.
DR PIR; S55124; S55124.
DR SGD; S0004789; MMT1.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR GO; GO:0006879; P:iron ion homeostasis; IGI.
DR InterPro; IPR002524; Cation_efflux.
DR Pfam; PF01545; Cation_efflux; 1.
DR TIGRFAMs; TIGR01297; CDF; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 194 214 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
SQ SEQUENCE 510 AA; 56209 MW; F3CC9A230FB5DB87 CRC64;

Query Match 79.3%; Score 23; DB 1; Length 510;
Best Local Similarity 37.5%; Pred. No. 4.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 137 HTHSHAAH 144

RESULT 3
BRC4_DROME STANDARD; PRT; 880 AA.
ID BRC4_DROME
AC Q24206; O46064; Q9W575;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Broad-complex core-protein isoform 6.
GN BR OR BR-C OR EG:17A9.1 OR EG:25D2.1 OR EG:123F11.1 OR
OS CG11491/CG11514.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND CHARACTERIZATION OF
RP ISOFORMS.
RC TISSUE=Imaginal disks, and Larva;
RX MEDLINE=96299417; PubMed=8660872;
RA Bayer C.A., Holley B., Fristrom J.W.;
RT "A switch in broad-complex zinc-finger isoform expression is regulated
RT posttranscriptionally during the metamorphosis of Drosophila imaginal
RT discs.";
RL Dev. Biol. 177:1-14(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,
RA Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster.";
RL Science 287:2220-2222(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [4]
RP CHARACTERIZATION OF ISOFORMS, AND MUTATIONAL ANALYSIS.
RX MEDLINE=97384928; PubMed=9242423;
RA Bayer C.A., von Kalm L., Fristrom J.W.;
RT "Relationships between protein isoforms and genetic functions
RT demonstrate functional redundancy at the Broad-Complex during
RT Drosophila metamorphosis.";
RL Dev. Biol. 187:267-282(1997).
CC -!- FUNCTION: BROAD-COMPLEX PROTEINS ARE REQUIRED FOR PUFFING AND
CC TRANSCRIPTION OF SALIVARY GLAND LATE GENES DURING METAMORPHOSIS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=6;
CC Name=6; Synonyms=BCORE-Z4;
CC IsoId=Q24206-1; Sequence=Displayed;
CC Name=1; Synonyms=BCORE-TNT1-Q1-Z1;
CC IsoId=Q01295-1; Sequence=External;
CC Name=2; Synonyms=BCORE-Q1-Z1;
CC IsoId=Q01295-2; Sequence=External;
CC Name=3; Synonyms=BCORE-Q2-Z1;
CC IsoId=Q01295-3; Sequence=External;
CC Name=4; Synonyms=BCORE-Z2;
CC IsoId=Q01295-4; Sequence=External;
CC Name=5; Synonyms=BCORE-NS-Z3;
CC IsoId=Q01295-5; Sequence=External;
CC -!- DEVELOPMENTAL STAGE: ACCUMULATES TO A HIGH LEVEL AT THE BEGINNING
CC OF THE ECDYSONE RESPONSE, DURING THE METAMORPHOSIS OF IMAGINAL
CC DISKS IN PUFF STAGE 1, AND ABRUPTLY DISAPPEARS AFTER SEVERAL
CC HOURS.
CC -!- INDUCTION: INDUCED AS A PRIMARY RESPONSE TO 20-HYDROXYECDYSONE IN
CC THIRD INSTAR LARVAL IMAGINAL DISKS.
CC -!- SIMILARITY: Contains 1 BTB/POZ domain.
CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC TO 619 AND 656 TO 694 DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U51585; AAB09760.1; ALT_FRAME.
CC EMBL; AL009146; CAA15627.1; -.
CC EMBL; AB003421; AAF45647.1; -.
CC TRANSFAC; T01480; -.
CC FlyBase; FBgn0000210; br.
CC InterPro; IPR000210; BTB_POZ.
CC InterPro; IPR007087; Znf_C2H2.
CC Pfam; PF00651; BTB; 1.
CC Pfam; PF00096; zf-C2H2; 2.
CC SMART; SM00225; BTB; 1.
CC SMART; SM00355; Znf_C2H2; 2.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; 2.
CC PROSITE; PS50097; BTB; 1.
CC Nuclear protein; DNA-binding; Developmental protein;
CC Zinc-finger; Metal-binding; Alternative splicing.
CC DOMAIN 32 97
CC ZN_FING 710 733 C2H2-TYPE 1.
CC ZN_FING 740 763 C2H2-TYPE 2.
CC DOMAIN 203 207 POLY-ALA.
CC DOMAIN 265 268 POLY-ASN.
CC DOMAIN 458 466 POLY-ASN.
CC DOMAIN 584 589 POLY-PRO.
CC DOMAIN 618 621 POLY-ALA.
CC DOMAIN 798 803 POLY-ALA.

```
FT DOMAIN 821 833 POLY-ALA.
FT DOMAIN 862 867 POLY-GLN.
FT CONFLICT 436 436 G -> D (IN REF. 1).
FT CONFLICT 621 621 MISSING (IN REF. 1).
FT CONFLICT 624 624 A -> R (IN REF. 1).
FT CONFLICT 661 662 AV -> L (IN REF. 1).
FT CONFLICT 668 678 MISSING (IN REF. 1).
FT CONFLICT 722 723 KL -> NV (IN REF. 1).
SQ SEQUENCE 880 AA; 92305 MW; 500C0A4A38663AAF CRC64;

Query Match 79.3%; Score 23; DB 1; Length 880;
Best Local Similarity 37.5%; Pred. No. 7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 625 HAHAAAH 632

RESULT 4
SYA_STRCO
ID SYA_STRCO STANDARD; PRT; 890 AA.
AC Q9KXP9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
GN ALAS OR SCO1501 OR SC9C5.25C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
SEQUENCE FROM N.A.
STRAIN=A3(2) / M145;
MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
SQ SEQUENCE 890 AA; 95786 MW; 05B2FD563D35F4DF CRC64;

Query Match 79.3%; Score 23; DB 1; Length 890;
Best Local Similarity 37.5%; Pred. No. 7.1e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 573 HSATHLTH 580

RESULT 5
GSRI_HUMAN
ID GSRI_HUMAN STANDARD; PRT; 1509 AA.
AC Q9NZM4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glioma tumor suppressor candidate region gene 1 protein.
GN GLTSCR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louis D.N., Jenkins R.B.;
RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
region.";
RL Genomics 64:44-50(2000).
CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
CC placenta, skeletal muscle, and pancreas, and at lower levels in
CC lung, liver, and kidney.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO YEAST YHR217C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z71614; CAA96274.1; -.
CC EMBL; Z71613; CAA96273.1; -.
DR PIR; S63324; S63324.
DR SGD; S0005282; YNL338W.
KW Hypothetical protein.
SQ SEQUENCE 52 AA; 5951 MW; C1E4066D43E057A1 CRC64;

Query Match 75.9%; Score 22; DB 1; Length 52;
Best Local Similarity 37.5%; Pred. No. 90;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 38 HTHHTHH 45

RESULT 7
Y034_TREPA
ID Y034_TREPA STANDARD; PRT; 316 AA.
AC O83077;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative periplasmic metal-binding protein TP0034 precursor.
GN TP0034.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -!- FUNCTION: PART OF AN ATP-DRIVEN TRANSPORT SYSTEM
CC TP0034/TP0035/TP0036 FOR A METAL. METAL-BINDING COMPONENT.
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL SOLUTE-BINDING PROTEIN FAMILY
CC 9.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; AE001188; AAC65029.1; -.
DR PIR; D71375; D71375.
DR TIGR; TP0034; -.
DR InterPro; IPR006128; Lipoprotein_4.
DR InterPro; IPR006127; SBP_bac_9.
DR Pfam; PF01297; SBP_bac_9; 1.
DR PRINTS; PR00690; ADHESNFAMILY.
DR PROSITE; PS00013; PROKAR LIPOPROTEIN; UNKNOWN 1.
KW Hypothetical protein; Transport; Periplasmic; Metal-binding; Signal;
KW Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 316 PUTATIVE PERIPLASMIC METAL-BINDING
FT PROTEIN TP0034.
SQ SEQUENCE 316 AA; 35433 MW; 16051C2199BC81AB CRC64;

Query Match 75.9%; Score 22; DB 1; Length 316;
Best Local Similarity 37.5%; Pred. No. 4.4e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 124 HTRGHTAH 131

RESULT 8
ADH1_BACST
ID ADH1_BACST STANDARD; PRT; 337 AA.
AC P12311;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase (EC 1.1.1.1) (ADH-T).
GN ADHT.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC STRAIN=NCA 1503;
RX MEDLINE=92138636; PubMed=1735726;
RA Sakoda H., Imanaka T.;
RT "Cloning and sequencing of the gene coding for alcohol dehydrogenase
RT of Bacillus stearothermophilus and rational shift of the optimum
RT pH.";
RL J. Bacteriol. 174:1397-1402(1992).
RN [2]
RP SEQUENCE OF 1-45.
RX MEDLINE=73229257; PubMed=4578954;
RA Bridgen J., Kolb E., Harris J.I.;
RT "Amino acid sequence homology in alcohol dehydrogenase.";
RL FEBS Lett. 33:1-3(1973).
RN [3]
RP SEQUENCE OF 34-54.
RX MEDLINE=79169263; PubMed=436831;
RA Jeck R., Woenckhaus C., Harris J.I., Runswick M.J.;
RT "Identification of the amino acid residue modified in Bacillus
RT stearothermophilus alcohol dehydrogenase by the NAD+ analogue 4-(3-
RT bromoacetylpyridinio)butyldiphosphoadenosine.";
RL Eur. J. Biochem. 93:57-64(1979).
RN [4]
RP SEQUENCE OF 1-37; 188-197; 247-263 AND 324-336.
RC STRAIN=NCA 1503;
RX MEDLINE=94325354; PubMed=8049268;
RA Robinson G.A., Bailey C.J., Dowds B.C.A.;
RT "Gene structure and amino acid sequences of alcohol dehydrogenases of
RT Bacillus stearothermophilus.";
RL Biochim. Biophys. Acta 1218:432-434(1994).
CC -!- FUNCTION: THERMOSTABLE NAD(+)-DEPENDENT ALCOHOL DEHYDROGENASE.
CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- ENZYME REGULATION: SUBSTRATE INHIBITION IS NOT OBSERVED WITH ANY
```



```
CC ALCOHOLS, AND THE ENZYME-NADH DISSOCIATION IS NOT CONSIDERED TO BE
CC A RATE-LIMITING STEP.
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90421; BAA14411.1; -.
CC PIR; A42654; A42654.
CC InterPro; IPR002328; ADH_zinc.
CC InterPro; IPR002085; Adh_zn_family.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC PROSITE; PS00059; ADH_ZINC; 1.
CC Oxidoreductase; Zinc; Metal-binding; NAD.
CC METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC METAL 61 61 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC METAL 92 92 ZINC 2 (BY SIMILARITY).
CC METAL 95 95 ZINC 2 (BY SIMILARITY).
CC METAL 98 98 ZINC 2 (BY SIMILARITY).
CC METAL 106 106 ZINC 2 (BY SIMILARITY).
CC METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC MUTAGEN 40 40 T->S: LITTLE DECREASE IN ACTIVITY.
CC MUTAGEN 43 43 H->R: HIGHER LEVEL OF ACTIVITY AT PH 9.
CC MUTAGEN 38 38 C->S: NO ACTIVITY.
CC MUTAGEN 40 40 T->A: NO ACTIVITY.
CC MUTAGEN 43 43 H->A: NO ACTIVITY.
CC CONFLICT 22 22 MISSING (IN REF. 2).
CC CONFLICT 33 33 MISSING (IN REF. 2).
CC CONFLICT 52 53 KP -> PK (IN REF. 3).
CC SEQUENCE 337 AA; 36100 MW; B9B35A80EE9B7A86 CRC64;

Query Match 75.9%; Score 22; DB 1; Length 337;
Best Local Similarity 37.5%; Pred. No. 4.6e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
DB 39 HTDLHAH 46

RESULT 9
ADH2_BACST STANDARD; PRT; 339 AA.
AC P42327;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase (EC 1.1.1.1) (ADH).
GN ADH.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.
RC STRAIN=DSM 2334;
RX MEDLINE=94325354; PubMed=8049268;
RA Robinson G.A., Bailey C.J., Dows B.C.A.;
RT "Gene structure and amino acid sequences of alcohol dehydrogenases of
RT Bacillus stearothermophilus.";
RL Biochim. Biophys. Acta 1218:432-434(1994).
CC -!- FUNCTION: ACTIVE WITH PRIMARY ALCOHOLS, INCLUDING METHANOL.
CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- ENZYME REGULATION: THE RATE-LIMITING STEP IS NADH RELEASE.
CC -!- CATABOLITE REPRESSION.
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
```

```
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z25544; CAA80989.1; -.
CC PIR; S47643; S47643.
CC InterPro; IPR002328; ADH_zinc.
CC InterPro; IPR002085; Adh_zn_family.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC PROSITE; PS00059; ADH_ZINC; 1.
CC Oxidoreductase; Zinc; Metal-binding; NAD.
CC METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC METAL 61 61 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC METAL 92 92 ZINC 2 (BY SIMILARITY).
CC METAL 95 95 ZINC 2 (BY SIMILARITY).
CC METAL 98 98 ZINC 2 (BY SIMILARITY).
CC METAL 106 106 ZINC 2 (BY SIMILARITY).
CC METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC SEQUENCE 339 AA; 36205 MW; 0EC33CE7287D7476 CRC64;

Query Match 75.9%; Score 22; DB 1; Length 339;
Best Local Similarity 37.5%; Pred. No. 4.6e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
DB 39 HTDLHAH 46

RESULT 10
ADH3_BACST STANDARD; PRT; 339 AA.
AC P42328;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase (EC 1.1.1.1) (ADH-HT).
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 12403 / LLD-R;
RX MEDLINE=94291628; PubMed=8020473;
RA Cannio R., Rossi M., Bartolucci S.;
RT "A few amino acid substitutions are responsible for the higher
RT thermostability of a novel NAD(+)-dependent bacillar alcohol
RT dehydrogenase.";
RL Eur. J. Biochem. 222:345-352(1994).
CC -!- FUNCTION: THERMOSTABLE AND THERMOPHILIC NAD(+)-DEPENDENT
CC ALCOHOL DEHYDROGENASE. BEARS MAINLY AN ETHANOL-DEHYDROGENASE
CC ACTIVITY.
CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z27089; CAA81612.1; -.
CC
```



```
DR PIR; S45605; S45605.
DR HSSP; P28304; IQOR.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD.
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 61 61 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 92 92 ZINC 2 (BY SIMILARITY).
FT METAL 95 95 ZINC 2 (BY SIMILARITY).
FT METAL 98 98 ZINC 2 (BY SIMILARITY).
FT METAL 106 106 ZINC 2 (BY SIMILARITY).
FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 339 AA; 36338 MW; AED17E4A34163430 CRC64;

Query Match 75.9%; Score 22; DB 1; Length 339;
Best Local Similarity 37.5%; Pred. No. 4.6e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 39 HTDLHAAH 46

RESULT 11
FILE_HUMAN
ID FILA_HUMAN STANDARD; PRT; 416 AA.
AC P20930;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Filaggrin precursor (Fragment).
GN FLG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89296901; PubMed=2740331;
RA McKinley-Grant L.J., Idler W.W., Bernstein I.A., Parry D.A.D.,
RA Cannizzaro L., Croce C.M., Huebner K., Lessin S.R., Steinert P.M.;
RT "Characterization of a cDNA clone encoding human filaggrin and
RT localization of the gene to chromosome region 1q21.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989).
RN [2]
RP CITRULLINATION.
RX MEDLINE=96374388; PubMed=8780679;
RA Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.;
RT "Preferential deimination of keratin K1 and filaggrin during the
RT terminal differentiation of human epidermis.";
RL Biochem. Biophys. Res. Commun. 225:712-719(1996).
CC -!- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC -!- PTM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC -!- PTM: Undergoes deimination of some arginine residues
CC (citrullination).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24355; AAA52454.1; -.
```

```
DR PIR; A32947; A32947.
DR Genew; HGNC:3748; FLG.
DR MIM; 135940; -.
DR GO; GO:0005882; C:intermediate filament; NAS.
DR GO; GO:0005198; F:structural molecule activity; NAS.
DR GO; GO:0007275; P:development; NAS.
DR InterPro; IPR003303; Filaggrin.
DR Pfam; PF03516; Filaggrin; 2.
DR PRINTS; PR00487; FILAGGRIN.
KW Phosphorylation; Citrullination; Developmental protein.
FT NON_TER 1 1
SQ SEQUENCE 416 AA; 44105 MW; DEEA3218BA043F32 CRC64;

Query Match 75.9%; Score 22; DB 1; Length 416;
Best Local Similarity 37.5%; Pred. No. 5.6e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | |
Db 99 HSGSHSH 106

RESULT 12
PFTB_PEA
ID PFTB_PEA STANDARD; PRT; 419 AA.
AC Q04903;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein farnesyltransferase beta subunit (EC 2.5.1.-) (CAAX
DE farnesyltransferase beta subunit) (RAS proteins prenyltransferase
DE beta) (Ftase-beta).
GN FTB.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Alaska; TISSUE=Seedling;
RX MEDLINE=94105305; PubMed=8278509;
RA Yang Z., Cramer C.L., Watson J.C.;
RT "Protein farnesyltransferase in plants. Molecular cloning and
RT expression of a homolog of the beta subunit from the garden pea.";
RL Plant Physiol. 101:667-674(1993).
CC -!- FUNCTION: CATALYZES THE TRANSFER OF A FARNESYL MOIETY FROM
CC FARNESYL PYROPHOSPHATE TO A CYSTEINE AT THE FOURTH POSITION FROM
CC THE C-TERMINUS OF SEVERAL PROTEINS. THE BETA SUBUNIT IS
CC RESPONSIBLE FOR PEPTIDE-BINDING (BY SIMILARITY).
CC -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PROTEIN PRENYLTRANSFERASE BETA SUBUNIT
CC FAMILY.
CC -!- SIMILARITY: Contains 5 PFTB repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L08664; AAA33649.1; -.
DR PIR; JQ2254; JQ2254.
DR HSSP; Q02293; 1FT1.
DR InterPro; IPR001330; Prenyltrans.
DR Pfam; PF00432; Prenyltrans; 5.
KW Transferase; Prenyltransferase; Repeat; Zinc.
FT REPEAT 68 109 PFTB 1.
FT REPEAT 119 160 PFTB 2.
```

```
FT REPEAT      167      208      PFTB 3.
FT REPEAT      215      256      PFTB 4.
FT REPEAT      329      371      PFTB 5.
FT METAL       241      241      ZINC (BY SIMILARITY).
FT METAL       243      243      ZINC (BY SIMILARITY).
FT METAL       359      359      ZINC (BY SIMILARITY).
SQ SEQUENCE    419 AA; 46793 MW; 4F040E0094277D7C CRC64;

  Query Match      75.9%; Score 22; DB 1; Length 419;
  Best Local Similarity 37.5%; Pred. No. 5.6e+02;
  Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 HXXXHXXH 8
Db      291 HATSHIRH 298

RESULT 13
PROA_NEIMB
ID PROA_NEIMB STANDARD; PRT; 420 AA.
AC Q9JZG3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase) (GSA dehydrogenase).
DE PROA OR NMB1068.
GN Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.";
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: CATALYZES THE NADPH DEPENDENT REDUCTION OF L-GAMMA-GLUTAMYL 5-PHOSPHATE INTO L-GLUTAMATE 5-SEMIALDEHYDE AND PHOSPHATE. THE PRODUCT SPONTANEOUSLY UNDERGOES CYCLIZATION TO FORM 1-PYRROLINE-5-CARBOXYLATE.
CC -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate + NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC -!- PATHWAY: Proline biosynthesis; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; AE002457; AAF62324.1; -.
DR TIGR; NMB1068; -.
DR HAMAP; MF_00412; -.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR TIGRFAMs; TIGR00407; proA; 1.
DR PROSITE; PS01223; PROA; 1.
KW Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ SEQUENCE 420 AA; 45241 MW; A5D96CEDE50E87A2 CRC64;

DR PROSITE; PS01223; PROA; 1.
KW Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ SEQUENCE 420 AA; 45256 MW; 009996E9CF6B1118 CRC64;

  Query Match      75.9%; Score 22; DB 1; Length 420;
  Best Local Similarity 37.5%; Pred. No. 5.6e+02;
  Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 HXXXHXXH 8
Db      335 HIETHSTH 342

RESULT 14
PROA_STRPN
ID PROA_STRPN STANDARD; PRT; 420 AA.
AC Q97R94;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase) (GSA dehydrogenase).
DE PROA OR SP0932.
GN Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D., Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J., Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D., Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";
RL Science 293:498-506(2001).
CC -!- FUNCTION: Catalyzes the NADPH dependent reduction of L-gamma-glutamyl 5-phosphate into L-glutamate 5-semialdehyde and phosphate. The product spontaneously undergoes cyclization to form 1-pyrroline-5-carboxylate.
CC -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate + NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC -!- PATHWAY: Proline biosynthesis; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; AE007398; AAK75056.1; -.
DR PIR; G95107; G95107.
DR TIGR; SP0932; -.
DR HAMAP; MF_00412; -.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR000965; Gglut_pp_reduct.
DR Pfam; PF00171; aldedh; 1.
DR TIGRFAMs; TIGR00407; proA; 1.
DR PROSITE; PS01223; PROA; 1.
KW Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ SEQUENCE 420 AA; 45241 MW; A5D96CEDE50E87A2 CRC64;
```

Query Match 75.9%; Score 22; DB 1; Length 420;
Best Local Similarity 37.5%; Pred. No. 5.6e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 335 HIESHSTH 342

RESULT 15
FKH2_XENLA
ID_FKH2_XENLA STANDARD; PRT; 427 AA.
AC P32315;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE XFKH2 protein.
GN XFKH2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94074768; PubMed=8253274;
RA Bolce M.E., Hemmati-Brivlanlou A., Harland R.M.;
RT "XFKH2, a Xenopus HNF-3 alpha homologue, exhibits both
RT activin-inducible and autonomous phases of expression in early
RT embryos.";
RL Dev. Biol. 160:413-423 (1993).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: PRESENT IN THE VEGETAL POLE AND MARGINAL ZONE
CC BUT NOT THE ANIMAL POLE OF GASTRULAE AND IN EQUAL LEVELS IN THE
CC DORSAL AND VENTRAL HALVES OF BOTH GASTRULAE AND NEURULAE.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M93658; AAA17050.1; -.
DR PIR; I51580; I51580.
DR HSSP; Q63245; 2HFH.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein.
FT DNA_BIND 156 247 FORK-HEAD.
SQ SEQUENCE 427 AA; 46572 MW; 2D29A42AF960730C CRC64;

Query Match 75.9%; Score 22; DB 1; Length 427;
Best Local Similarity 37.5%; Pred. No. 5.7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 333 HSLAHETH 340

CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X62088; CAA43998.1; --
DR EMBL; X53180; CAA37321.1; --
DR PIR; JH0256; JH0256.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; --
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0 0
FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 212 212 BY SIMILARITY.
FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 411 425 INTERCHAIN (PROBABLE).
FT CONFLICT 229 229 K -> M (IN REF. 2).
SQ SEQUENCE 1250 AA; 143265 MW; 8171B5B2C2312857 CRC64;

Query Match 78.3%; Score 36; DB 1; Length 1250;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 211 HELIHSFH 218

RESULT 8
RECR_LISIN
ID RECR LISIN STANDARD; PRT; 198 AA.
AC Q927D9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Recombination protein recr.
GN RECR OR LIN2850.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";

RL Science 294:849-852(2001).
CC -!- FUNCTION: May play a role in DNA repair. It seems to be involved
CC in an recBC-independent recombinational process of DNA repair. It
CC may act with recF and recO (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL596173; CAC98076.1; --
DR PIR; AD1788; AD1788.
DR ListList; LIN02850; --
DR HAMAP; MF_00017; --; 1.
DR InterPro; IPR003583; HHH 1.
DR InterPro; IPR000093; RecR.
DR InterPro; IPR006171; Toprim_dom.
DR InterPro; IPR006154; Toprim_sub.
DR Pfam; PF02132; RecR; 1.
DR Pfam; PF01751; Toprim; 1.
DR SMART; SM00278; HhH1; 1.
DR SMART; SM00493; TOPRIM; 1.
DR TIGRFAMS; TIGR00615; recR; 1.
DR PROSITE; PS01300; RECR; 1.
KW DNA repair; DNA recombination; Zinc-finger; Complete proteome.
FT ZN_FING 57 72 C4-TYPE (POTENTIAL).
SQ SEQUENCE 198 AA; 21996 MW; 8A82E1A16415DFFC CRC64;

Query Match 73.9%; Score 34; DB 1; Length 198;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 101 HGLYHVLH 108

RESULT 9
RECR LISMO
ID RECR LISMO STANDARD; PRT; 198 AA.
AC Q8Y3X7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Recombination protein recr.
GN RECR OR LMO2702.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";

-!- FUNCTION: May play a role in DNA repair. It seems to be involved
in an recBC-independent recombinational process of DNA repair. It
may act with recF and recO (By similarity).

CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL591984; CAD00915.1; -.
DR PIR; AE1412; AE1412.
DR Listlist; LMO02702; -.
DR HAMAP; MF_00017; -; 1.
DR InterPro; IPR003583; HHH 1.
DR InterPro; IPR000093; RecR.
DR InterPro; IPR006171; Toprim_dom.
DR InterPro; IPR006154; Toprim_sub.
DR Pfam; PF02132; RecR; 1.
DR Pfam; PF01751; Toprim; 1.
DR SMART; SM00278; HhH1; 1.
DR SMART; SM00493; TOPRIM; 1.
DR TIGRFAMS; TIGR00615; recR; 1.
DR PROSITE; PS01300; RecR; 1.
KW DNA repair; DNA recombination; Zinc-finger; Complete proteome.
FT ZN FING 57 72 C4-TYPE (POTENTIAL).
SQ SEQUENCE 198 AA; 21934 MW; E542E27BC3D05036 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 198;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 101 HGLYHVLH 108

RESULT 10
RECR STRPN STANDARD; PRT; 198 AA.
AC Q9ZHC4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Recombination protein recR.
GN RECR OR RECM OR SP1672 OR SPR1516.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G54 / Type 19F;
RX MEDLINE=99061199; PubMed=9846742;
RA Massidda O., Anderluzzi D., Friedli L., Feger G.;
RT "Unconventional organization of the division and cell wall gene
RT cluster of Streptococcus pneumoniae.";
RL Microbiology 144:3069-3078(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=339, and PN94-661;
RX MEDLINE=20073037; PubMed=10605111;
RA Enright M.C., Spratt B.G.;
RT "Extensive variation in the ddl gene of penicillin-resistant
RT Streptococcus pneumoniae results from a hitchhiking effect driven by
RT the penicillin-binding protein 2b gene.";
RL Mol. Biol. Evol. 16:1687-1695(1999).
RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,

RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
CC -!- FUNCTION: MAY PLAY A ROLE IN DNA REPAIR. IT SEEMS TO BE INVOLVED
CC IN AN RECBC-INDEPENDENT RECOMBINATIONAL PROCESS OF DNA REPAIR. IT
CC MAY ACT WITH RECF AND RECO (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF068901; AAC95434.1; -.
DR EMBL; AJ243056; CAB64474.1; -.
DR EMBL; AJ243057; CAB64478.1; -.
DR EMBL; AE007460; AAK75751.1; -.
DR EMBL; AE008520; AAL00320.1; -.
DR PIR; C98061; C98061.
DR PIR; F95194; F95194.
DR TIGR; SP1672; -.
DR HAMAP; MF_00017; -; 1.
DR InterPro; IPR000093; RecR.
DR InterPro; IPR006171; Toprim_dom.
DR InterPro; IPR006154; Toprim_sub.
DR Pfam; PF02132; RecR; 1.
DR Pfam; PF01751; Toprim; 1.
DR SMART; SM00493; TOPRIM; 1.
DR TIGRFAMS; TIGR00615; recR; 1.
DR PROSITE; PS01300; RecR; 1.
KW DNA repair; DNA recombination; Zinc-finger; Complete proteome.
FT ZN FING 57 72 C4-TYPE (POTENTIAL).
SQ SEQUENCE 198 AA; 21689 MW; FC6F0E98E3933752 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 198;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 101 HGLYHVLH 108

RESULT 11
SODC HAEDU STANDARD; PRT; 199 AA.
ID SODC HAEDU
AC Q59452; Q59449; Q59453;
DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
GN SODC.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000;
RX MEDLINE=97288949; PubMed=9143881;
RA Langford P.R., Kroll J.S.;
RT "Distribution, cloning, characterisation and mutagenesis of sodC, the
RT gene encoding copper/zinc superoxide dismutase, a potential
RT determinant of virulence, in Haemophilus ducreyi.";
RL FEMS Immunol. Med. Microbiol. 17:235-242(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=35000;
RX MEDLINE=97149276; PubMed=8996084;
RA Stevens M.K., Hassett D.J., Radolf J.D., Hansen E.J.;
RT "Cloning and sequencing of the gene encoding the Cu,Zn-superoxide
RT dismutase of Haemophilus ducreyi.";
RL Gene 183:35-40(1996).
RN [3]
RP SEQUENCE OF 100-186 FROM N.A.
RC STRAIN=35000;
RX MEDLINE=96118708; PubMed=7496539;
RA Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;
RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
RT from the eukaryotic enzyme, and not so rare after all!";
RL Microbiology 141:2271-2279(1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems. May play a role
CC in the interactive biology of organisms with their hosts and so
CC contribute to their capacity to cause disease.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X98737; CAA67289.1; -.
DR EMBL; U47664; AAB41293.1; -.
DR EMBL; X83125; CAA58206.1; -.
DR PIR; JC5718; JC5718.
DR HSSP; P24702; 2APS.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; sodcu; 1.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00087; SOD_CU_ZN_1; 1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW Antioxidant; Oxidoreductase; Metal-binding; Copper; Zinc; Periplasmic;
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 199 SUPEROXIDE DISMUTASE [CU-ZN].
FT METAL 92 92 COPPER (BY SIMILARITY).
FT METAL 94 94 COPPER (BY SIMILARITY).
FT METAL 117 117 COPPER AND ZINC (BY SIMILARITY).
FT METAL 126 126 ZINC (BY SIMILARITY).
FT METAL 135 135 ZINC (BY SIMILARITY).
FT METAL 138 138 ZINC (BY SIMILARITY).
FT METAL 173 173 COPPER (BY SIMILARITY).

FT DISULFID 99 195 BY SIMILARITY.
SQ SEQUENCE 199 AA; 21402 MW; 841D3210AB2BC06C CRC64;

Query Match 73.9%; Score 34; DB 1; Length 199;
Best Local Similarity 75.0%; Pred.No. 13;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
||| |
Db 82 HDLAHGLH 89

RESULT 12
EX7L VIBCH STANDARD; PRT; 446 AA.
ID EX7L VIBCH
AC Q9KTW4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR VC0766.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Winn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -!- FUNCTION: BIDIRECTIONALLY DEGRADES SINGLE-STRANDED DNA INTO LARGE
CC ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
CC INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE004162; AAF93931.1; -.
DR PIR; B82282; B82282.
DR TIGR; VC0766; -.
DR HAMAP; MF_00378; -; 1.
DR InterPro; IPR003753; Exonuc_VII_L.
DR InterPro; IPR004365; tRNA_anti.
DR Pfam; PF02601; Exonuc_VII_L; 1.
DR Pfam; PF01336; tRNA_anti; 1.
DR TIGRFAMs; TIGR00237; xsea; 1.
KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 446 AA; 50542 MW; AAEL7369636A4BC9 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 446;
Best Local Similarity 71.4%; Pred.No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 HDLIHVL 7
      ||::|||
Db     155 HDILHVL 161

RESULT 13
EX7L_SALTI
ID_EX7L_SALTI STANDARD; PRT; 449 AA.
AC Q8Z4Q1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR STY2753 OR T0345.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltham T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627275; CAD02714.1; -.
CC DR EMBL; AE016835; AAO68065.1; -.
CC DR HAMAP; MF_00378; -.
CC DR InterPro; IPR003753; Exonuc VII_L.
CC DR InterPro; IPR004365; tRNA_anti.
CC DR Pfam; PF02601; Exonuc VII_L; 1.
CC DR Pfam; PF01336; tRNA_anti; 1.
CC DR TIGRFAMs; TIGR00237; xsea; 1.
CC KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
CC SEQUENCE 449 AA; 50720 MW; 511957DEC878F5D2 CRC64;
CC -----
QY      1 HDLIHVL 7
      ||::|||
Db     154 HDILHVL 160

Query Match 73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
EX7L_SALTY
ID_EX7L_SALTY STANDARD; PRT; 449 AA.
AC Q8ZN58;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR STM2512.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008813; AAL21406.1; -.
CC DR StyGene; SG????; xsea.
CC DR HAMAP; MF_00378; -.
CC DR InterPro; IPR003753; Exonuc VII_L.
CC DR InterPro; IPR004365; tRNA_anti.
CC DR Pfam; PF02601; Exonuc VII_L; 1.
CC DR Pfam; PF01336; tRNA_anti; 1.
CC DR TIGRFAMs; TIGR00237; xsea; 1.
CC KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
CC SEQUENCE 449 AA; 50613 MW; 85356CE8560E161E CRC64;
CC -----
QY      1 HDLIHVL 7
      ||::|||
Db     154 HDILHVL 160

Query Match 73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:32:27 ; Search time 28 Seconds
(without alignments)
73.729 Million cell updates/sec

Title: US-10-064-903-2
Perfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	451	2 Q9R631	Q9r631 clostridium
2	42	91.3	1268	2 Q45851	Q45851 clostridium
3	42	91.3	1291	2 Q9ZAJ8	Q9zaj8 clostridium
4	42	91.3	1291	2 Q93G71	Q93g71 clostridium
5	42	91.3	1291	2 Q933K0	Q933k0 clostridium
6	42	91.3	1291	2 Q08077	Q08077 clostridium
7	42	91.3	1291	2 Q8GR96	Q8gr96 clostridium
8	42	91.3	1310	2 Q93N27	Q93n27 clostridium
9	39	84.8	397	16 Q92ML7	Q92ml7 rhizobium m
10	38	82.6	925	16 Q9JS16	Q9jsi6 chlamydia p
11	38	82.6	925	16 Q9Z8P5	Q9z8p5 chlamydia p
12	38	82.6	1278	2 Q57236	Q57236 clostridium
13	38	82.6	1280	2 Q9ZAJ5	Q9zaj5 clostridium
14	37	80.4	349	5 Q8IJV0	Q8ijv0 plasmodium
15	37	80.4	707	3 Q8X008	Q8x008 neurospora
16	36	78.3	105	10 Q8VXL7	Q8vxl7 fagus sylv

17	36	78.3	232	16	Q8X829	Q8x829 escherichia
18	36	78.3	237	13	Q8AWC9	Q8awc9 cyprinus ca
19	36	78.3	241	10	Q8VXL6	Q8vxl6 fagus sylv
20	36	78.3	489	9	Q9AZH2	Q9azh2 bacterioph
21	36	78.3	489	16	Q9CEA2	Q9cea2 lactococcus
22	36	78.3	773	10	Q9SHP2	Q9shp2 arabidopsis
23	36	78.3	1251	2	Q9K395	Q9k395 clostridium
24	36	78.3	1252	2	Q8KZM3	Q8kzm3 clostridium
25	36	78.3	1255	2	Q9FAR6	Q9far6 clostridium
26	35	76.1	312	16	Q26068	Q26068 helicobacte
27	35	76.1	312	16	Q9ZJ59	Q9zj59 helicobacte
28	35	76.1	426	5	Q8MYP4	Q8myp4 caenorhabdi
29	35	76.1	431	5	Q9XTZ9	Q9xtz9 caenorhabdi
30	35	76.1	500	16	Q9RVQ8	Q9rvq8 deinococcus
31	35	76.1	679	16	Q97SL8	Q97sl8 streptococc
32	35	76.1	737	16	Q8DR83	Q8dr83 streptococc
33	34	73.9	129	3	Q8TFT7	Q8tft7 ustilago vi
34	34	73.9	198	2	Q9RCP8	Q9rcp8 streptococc
35	34	73.9	198	2	Q9RCQ5	Q9rcq5 streptococc
36	34	73.9	198	2	Q9RCR0	Q9rcr0 streptococc
37	34	73.9	198	2	Q9R2M1	Q9r2m1 streptococc
38	34	73.9	198	2	Q9RCQ2	Q9rcq2 streptococc
39	34	73.9	199	16	Q8DV99	Q8dv99 streptococc
40	34	73.9	222	16	Q8EE87	Q8ee87 shewanella
41	34	73.9	225	5	Q8I1V8	Q8i1v8 plasmodium
42	34	73.9	261	5	Q8IFQ1	Q8ifq1 plasmodium
43	34	73.9	299	17	Q96YN4	Q96yn4 sulfolobus
44	34	73.9	385	10	Q8H3N0	Q8h3n0 oryza sativ
45	34	73.9	458	16	Q8FF64	Q8ff64 escherichia

ALIGNMENTS

RESULT 1

Q9R631 PRELIMINARY; PRT; 451 AA.
ID Q9R631
AC Q9R631;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type B light chain, BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92340509; PubMed=1634516;
RA Kurazono H., Mochida S., Binz T., Eisel U., Quanz M., Grebenstein O.,
RA Wernars K., Poulain B., Tauc L., Niemann H.;
RT "Minimal essential domains specifying toxicity of the light chains of
RT tetanus toxin and botulinum neurotoxin type A.";
RL J. Biol. Chem. 267:14721-14729(1992).
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 451 AA; 51943 MW; 6C79FD488653EA71 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 451;
Best Local Similarity 87.5%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8

Db 230 HELIHVLH 237

RESULT 2

```
Q45851
ID Q45851 PRELIMINARY; PRT; 1268 AA.
AC Q45851;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type F.
GN BONT /F.
OS Clostridium baratii.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93252228; PubMed=8486245;
RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA Richardson P.T.;
RT "Nucleotide sequence of the gene coding for Clostridium baratii type F
RT neurotoxin: Comparison with other clostridial neurotoxins.";
RL FEMS Microbiol. Lett. 108:175-182(1993).
DR EMBL; X68262; CAA48329.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1268;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 219 HELIHLVH 226

RESULT 3
Q9ZAJ8 Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bont protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13630; CAA73968.1; -.
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150840 MW; E4D3B0E46AB2E735 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 HDLIHVLH 8
Db 230 HELIHLVH 237

RESULT 4
Q93G71 Q93G71 PRELIMINARY; PRT; 1291 AA.
AC Q93G71;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1436;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AF295926; AAK97132.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 230 HELIHLVH 237

RESULT 5
Q933K0 Q933K0 PRELIMINARY; PRT; 1291 AA.
AC Q933K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=593, and 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin.
KW Neurotoxin.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FA5A151 CRC64;
```

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 6

Q08077 ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE BONT/B.
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT Clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 7

Q08GR96 ID Q08GR96 PRELIMINARY; PRT; 1291 AA.
AC Q08GR96;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin.
GN BONTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RA Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA Nakamura S., Mukamoto M., Kozaki S.;
RT "Clostridium botulinum type B neurotoxin associated with infant
RT botulism.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB084152; BAC22064.1; -.
SQ SEQUENCE 1291 AA; 150574 MW; 0227CAEF4F58504D CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 8

Q93N27 ID Q93N27 PRELIMINARY; PRT; 1310 AA.
AC Q93N27;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tetanus toxin (Fragment).
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RA Shumin Z., Dianliang L.;
RT "Cloning and sequence analysis of tetanus toxin gene.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF389424; AAK72964.2; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1
FT NON_TER 1310 1310
SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1310;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 234 HELIHLVH 241

RESULT 9

Q92ML7 ID Q92ML7 PRELIMINARY; PRT; 397 AA.
AC Q92ML7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative deaminase OR deamidase protein.
GN R02596 OR SMC02420.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Puxnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";

```
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591791; CAC47175.1; -.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR001130; TatD_DNase.
DR Pfam; PF01979; Amidohydro_1; 1.
DR PROSITE; PS01137; TATD_1; 1.
KW Complete proteome.
SQ SEQUENCE 397 AA; 43054 MW; B7D5F69C4999CHE02 CRC64;

Query Match      84.8%; Score 39; DB 16; Length 397;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 326 HDLEHVLH 333

RESULT 10
Q9JS16 PRELIMINARY; PRT; 925 AA.
AC Q9JS16;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CT234 hypothetical protein.
GN CPJ0293 OR CP0465.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AF002208; AAF38302.1; -.
DR EMBL; AP002546; BAA98503.1; -.
DR TIGR; CP0465; -.
SQ SEQUENCE 925 AA; 105601 MW; 61E8941E7C8FD620 CRC64;

Query Match      82.6%; Score 38; DB 16; Length 925;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 538 HDLHITH 545

RESULT 11
Q9Z8P5 PRELIMINARY; PRT; 925 AA.
ID Q9Z8P5
AC Q9Z8P5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CT234 hypothetical protein.

RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591791; CAC47175.1; -.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR001130; TatD_DNase.
DR Pfam; PF01979; Amidohydro_1; 1.
DR PROSITE; PS01137; TATD_1; 1.
KW Complete proteome.
SQ SEQUENCE 397 AA; 43054 MW; B7D5F69C4999CHE02 CRC64;

Query Match      84.8%; Score 39; DB 16; Length 397;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 326 HDLEHVLH 333

RESULT 10
Q9JS16 PRELIMINARY; PRT; 925 AA.
AC Q9JS16;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CT234 hypothetical protein.
GN CPJ0293 OR CP0465.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AF002208; AAF38302.1; -.
DR EMBL; AP002546; BAA98503.1; -.
DR TIGR; CP0465; -.
SQ SEQUENCE 925 AA; 105601 MW; 61E8941E7C8FD620 CRC64;

Query Match      82.6%; Score 38; DB 16; Length 925;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 538 HDLHITH 545

RESULT 11
Q9Z8P5 PRELIMINARY; PRT; 925 AA.
ID Q9Z8P5
AC Q9Z8P5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CT234 hypothetical protein.

CPN0293.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
DR EMBL; AE001614; AAD18442.1; -.
KW Complete proteome.
SQ SEQUENCE 925 AA; 105615 MW; 98E6098E7C8FD37D CRC64;

Query Match      82.6%; Score 38; DB 16; Length 925;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 538 HDLHITH 545

RESULT 12
Q57236 PRELIMINARY; PRT; 1278 AA.
ID Q57236
AC Q57236; Q45863;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Botulinum neurotoxin type F (BONT/F protein).
GN BONT/F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10281;
RA Hutson R.A., Collins M.D.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE OF 635-1000 FROM N.A.
RC STRAIN=NCTC 1028;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=LANGELAND;
RX MEDLINE=98404102; PubMed=9732534;
RA East A.K., Bhandari M., Hielm S., Collins M.D.;
RT "Analysis of the botulinum neurotoxin type F gene clusters in
RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT barati.";
RL Curr. Microbiol. 37:262-268(1998).
DR EMBL; X81714; CAA57358.1; -.
DR EMBL; L35496; AAA23210.1; -.
DR EMBL; X70821; CAA50152.1; -.
DR EMBL; X99064; CAA67512.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn MTpeptidse.
DR Pfam; PF01742; Peptidase_M27; 1.
```



```
DR PRINTS; PR00760; BONT0XILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin.
SQ SEQUENCE 1278 AA; 147073 MW; 1B431D6918 CRC64;

Query Match
Best Local Similarity 82.6%; Score 38; DB 2; Length 1278;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 227 HELIHALH 234

RESULT 13
Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.
AC Q9ZAJ5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bont protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
complex in a strain of clostridium botulinum producing type B & F
neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13631; CAA73972.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTPeptidse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONT0XILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1280 AA; 147487 MW; D0F748976EBC222C CRC64;

Query Match
Best Local Similarity 82.6%; Score 38; DB 2; Length 1280;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 227 HELIHALH 234

RESULT 14
Q8IJV0 PRELIMINARY; PRT; 349 AA.
AC Q8IJV0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF10_0092.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
```

```
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AB014830; AAN35290.1; -.
KW Hypothetical protein.
SQ SEQUENCE 349 AA; 40775 MW; 9D7C20FEACED4464 CRC64;

Query Match
Best Local Similarity 80.4%; Score 37; DB 5; Length 349;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 96 HELVHIVH 103

RESULT 15
Q8X008 PRELIMINARY; PRT; 707 AA.
AC Q8X008;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Related to hydroxyproline-rich glycoprotein.
GN B23H20.050.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL669988; CAD21077.1; -.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 707 AA; 77817 MW; C49BC3C1A18D83F5 CRC64;

Query Match
Best Local Similarity 80.4%; Score 37; DB 3; Length 707;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 158 HDLLYLH 165

Search completed: November 12, 2003, 09:36:51
Job time : 31 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 12, 2003, 09:32:27 ; Search time 28 Seconds
(without alignments)
73.729 Million cell updates/sec

Title: US-10-064-903-1
Perfect score: 29
Sequence: 1 HXXXHXXH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	79.3	119	16	Q92K02 rhizobium m
2	23	79.3	152	17	Q9YE47 aeropyrum p
3	23	79.3	177	5	Q9XWQ9 caenorhabdi
4	23	79.3	204	2	Q9WX65 acetobacter
5	23	79.3	240	16	Q9PFU7 Q9pfu7 xylella fas
6	23	79.3	324	2	Q9F502 anabaena va
7	23	79.3	327	2	Q9FD42 anabaena sp
8	23	79.3	327	16	Q8YU45 anabaena sp
9	23	79.3	341	5	Q8T468 drosophila
10	23	79.3	341	5	Q8MY62 ciona savig
11	23	79.3	342	5	Q18888 caenorhabdi
12	23	79.3	382	16	O54173 streptomyce
13	23	79.3	402	2	Q9KX27 oligotropha
14	23	79.3	412	13	Q98U15 Q98ui5 lampetra ja
15	23	79.3	466	10	Q947K4 brassica na
16	23	79.3	472	5	Q09657 caenorhabdi

17	23	79.3	534	5	Q8MTV4 Q8mtv4 plasmodium
18	23	79.3	536	5	Q8I3N3 Q8i3n3 plasmodium
19	23	79.3	596	16	Q9L1Q1 Q9l1q1 streptomyce
20	23	79.3	606	16	O31566 bacillus su
21	23	79.3	610	2	O52961 bacillus su
22	23	79.3	635	4	Q13476 Q13476 homo sapien
23	23	79.3	676	5	Q95WV0 Q95wv0 drosophila
24	23	79.3	677	6	Q28256 Q28256 canis famil
25	23	79.3	690	5	Q8IM21 Q8im21 plasmodium
26	23	79.3	696	5	Q9VDP3 Q9vdp3 drosophila
27	23	79.3	713	16	Q8XR50 Q8xr50 ralstonia s
28	23	79.3	826	10	Q9SCU4 Q9scu4 arabidopsis
29	23	79.3	826	10	Q9SCU3 Q9scu3 arabidopsis
30	23	79.3	899	3	Q8NIZ0 Q8niz0 neurospora
31	23	79.3	1031	5	Q8MT64 Q8mt64 drosophila
32	23	79.3	1059	5	Q9VZ52 Q9vz52 drosophila
33	23	79.3	1182	5	Q9VXL1 Q9vxl1 drosophila
34	23	79.3	1226	5	Q9V4U3 Q9v4u3 drosophila
35	23	79.3	1359	5	Q9VX26 Q9vx26 drosophila
36	23	79.3	3036	4	Q8TDJ6 Q8tdj6 homo sapien
37	23	79.3	3469	5	Q9U4I2 Q9u4i2 drosophila
38	23	79.3	3604	5	Q9VYK0 Q9vyk0 drosophila
39	23	79.3	4360	3	Q9UVN5 Q9uvn5 alternaria
40	22	75.9	56	3	Q8TGJ7 Q8tgj7 saccharomyc
41	22	75.9	57	16	Q8PE57 Q8pe57 xanthomonas
42	22	75.9	61	16	Q8ZE32 Q8ze32 yersinia pe
43	22	75.9	77	12	Q8QN71 Q8qn71 ectocarpus
44	22	75.9	93	7	Q9GJ30 Q9gj30 homo sapien
45	22	75.9	93	7	Q9GJ32 Q9gj32 homo sapien

ALIGNMENTS

RESULT 1
Q92K02 ID Q92K02 PRELIMINARY; PRT; 119 AA.
AC Q92K02;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical transmembrane protein SMC01986.
GN R02520 OR SMC01986.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591791; CAC47099.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 119 AA; 13504 MW; CFFA1042CA34D6A9 CRC64;

Query Match 79.3%; Score 23; DB 16; Length 119;
Best Local Similarity 37.5%; Pred. No. 8e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 11 HAAAHTEH 18

RESULT 2

```
Q9YE47
ID Q9YE47 PRELIMINARY; PRT; 152 AA.
AC Q9YE47;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein APE0723.
GN APE0723.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79699.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 152 AA; 16298 MW; EE300BACCBA468F CRC64;

Query Match 79.3%; Score 23; DB 17; Length 152;
Best Local Similarity 37.5%; Pred. No. 9.8e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 15 HSTTHAAH 22

RESULT 3
Q9XWQ9 PRELIMINARY; PRT; 177 AA.
AC Q9XWQ9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Y11D7A.1 protein.
GN Y11D7A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Steward C.A.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL032632; CAA21589.1; -.
DR WormPep; Y11D7A.1; CE19027.
DR InterPro; IPR001220; Lectin legB.
DR PROSITE; PS00307; LECTIN LEGUME BETA; 1.
SQ SEQUENCE 177 AA; 19422 MW; 349B9DF2D33D17F1 CRC64;

Query Match 79.3%; Score 23; DB 5; Length 177;
Best Local Similarity 37.5%; Pred. No. 1.1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
```

```
Db 140 HTTVHSH 147

RESULT 4
Q9WX65 PRELIMINARY; PRT; 204 AA.
ID Q9WX65;
AC Q9WX65;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE BGL protein (Fragment).
GN BGL.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Gluconacetobacter.
OX NCBI_TaxID=28448;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM7664;
RA Umeda Y., Hirano A., Hon-nami K., Kunito S., Akiyama H., Onizuka T.,
RA Ikeuchi M., Inoue Y.;
RT "Conversion of CO2 into cellulose by gene manipulation of microalgae:
RT cloning of cellulose synthase genes from Acetobacter xylinum.";
RL (In) Inui T., Anpo M., Izui K., Yanagida S., Yanaguchi T. (eds.);
RL Advances in chemical conversions for mitigating carbon dioxide,
RL pp.114:653-656, Elsevier Science, Amsterdam (1998).
DR EMBL; AB015802; BAA77589.1; -.
DR InterPro; IPR001764; Glyco_hydro_3N.
DR Pfam; PF00933; Glyco_hydro_3; 1.
DR PRINTS; PR00133; GLHYDRLASE3.
FT NON TER 204
SQ SEQUENCE 204 AA; 21146 MW; 2CD1050D8E2E720F CRC64;

Query Match 79.3%; Score 23; DB 2; Length 204;
Best Local Similarity 37.5%; Pred. No. 1.2e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 21 HDAHAHAH 28

RESULT 5
Q9PFU7 PRELIMINARY; PRT; 240 AA.
ID Q9PFU7;
AC Q9PFU7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE GMP synthase.
GN XF0560.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
```

RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE003903; AAF83370.1; -.
DR InterPro; IPR000991; GATase_1.
DR Pfam; PF00117; GATase; 1.
DR PRINTS; PR00096; GATASE.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
KW Complete proteome.
SQ SEQUENCE 240 AA; 26350 MW; FF81E5BE1EBEEA35 CRC64;

Query Match 79.3%; Score 23; DB 16; Length 240;
Best Local Similarity 37.5%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 142 HFSAHATH 149

RESULT 6
Q9F502
ID Q9F502 PRELIMINARY; PRT; 324 AA.
AC Q9F502;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cytochrome c oxidase subunit 2 (EC 1.9.3.1).
GN COXB2.
OS Anabaena variabilis.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=1172;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC29413;
RA Pils D., Schmetterer G.;
RT "A second cytochrome c oxidase from the cyanobacterium Anabaena sp.
RT strain ATCC29413, up-regulated under nitrogen fixing conditions.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: COPPER A (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR EMBL; AJ296086; CAC12660.1; -.
DR InterPro; IPR001505; Copper_CuA.
DR Pfam; PF00116; COX2; 1.
DR Pfam; PF02790; COX2_TM; 1.
DR PRINTS; PR01166; CYCOXIDASEII.
DR ProDom; PD000131; Copper_CuA; 1.
DR PROSITE; PS00078; COX2; 1.
KW Copper; Oxidoreductase; Transmembrane.
SQ SEQUENCE 324 AA; 34739 MW; E16B6CC160899F72 CRC64;

Query Match 79.3%; Score 23; DB 2; Length 324;
Best Local Similarity 37.5%; Pred. No. 1.8e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 126 HASAHVAH 133

RESULT 7
Q9FD42
ID Q9FD42 PRELIMINARY; PRT; 327 AA.
AC Q9FD42;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cytochrome c oxidase subunit II.
GN CTAC.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7120;
RA Jones K.M., Buikema W.J., Haselkorn R.;
RT "Characterization of a heterocyst-specific cytochrome c oxidase operon
RT in Anabaena PCC7120.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: COPPER A (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR EMBL; AF291994; AAG01550.1; -.
DR InterPro; IPR001505; Copper_CuA.
DR InterPro; IPR002429; Cyt_c_ox_2.
DR Pfam; PF00116; COX2; 1.
DR Pfam; PF02790; COX2_TM; 1.
DR PRINTS; PR01166; CYCOXIDASEII.
DR ProDom; PD000131; Copper_CuA; 1.
DR PROSITE; PS00078; COX2; 1.
KW Copper; Oxidoreductase; Transmembrane.
SQ SEQUENCE 327 AA; 35003 MW; 92324730EB7A92F8 CRC64;

Query Match 79.3%; Score 23; DB 2; Length 327;
Best Local Similarity 37.5%; Pred. No. 1.8e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
| | | |
Db 126 HASAHVAH 133

RESULT 8
Q8YU45
ID Q8YU45 PRELIMINARY; PRT; 327 AA.
AC Q8YU45;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cytochrome c oxidase subunit II.
GN COXB OR ALR2514.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003589; BAB74213.1; -.
DR InterPro; IPR001505; Copper_CuA.
DR InterPro; IPR002429; Cyt_c_ox_2.
DR Pfam; PF00116; COX2; 1.
DR Pfam; PF02790; COX2_TM; 1.
DR PRINTS; PR01166; CYCOXIDASEII.
DR ProDom; PD000131; Copper_CuA; 1.

DR PROSITE; PS00078; COX2; 1.
KW Complete proteome.
SQ SEQUENCE 327 AA; 35006 MW; 3369DCBE31BA92E2 CRC64;

Query Match 79.3%; Score 23; DB 16; Length 327;
Best Local Similarity 37.5%; Pred. No. 1.8e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 126 HASARVAH 133

RESULT 9
Q8T468 PRELIMINARY; PRT; 341 AA.
AC Q8T468;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE AT13703p.
GN BCDNA:AT13703.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY089327; AAL90065.1; -;
DR FlyBase; FBgn0063732; BCDNA:AT13703.
SQ SEQUENCE 341 AA; 34098 MW; EE6EBF27B2EC618A CRC64;

Query Match 79.3%; Score 23; DB 5; Length 341;
Best Local Similarity 37.5%; Pred. No. 1.9e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 287 HSHSHAAH 294

RESULT 10
Q8MY62 PRELIMINARY; PRT; 341 AA.
AC Q8MY62;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Casein kinase I.
GN CS-CKI.
OS Ciona savignyi.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Cionidae; Ciona.
OX NCBI_TaxID=51511;
RN [1]
RP SEQUENCE FROM N.A.
RA Satou Y., Satoh N.;
RT "Ciona savignyi maternal gene."
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB078412; BAC05520.1; -;
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.

DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 341 AA; 39427 MW; 9BA8B5348512C185 CRC64;

Query Match 79.3%; Score 23; DB 5; Length 341;
Best Local Similarity 37.5%; Pred. No. 1.9e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 321 HTTQHASH 328

RESULT 11
Q18888 PRELIMINARY; PRT; 342 AA.
AC Q18888;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 39.5 kDa protein.
GN C56C10.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton L.;
RT "The sequence of C. elegans cosmid C56C10.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U29488; AAA68778.1; -;
DR WormPep; C56C10.10; CE02564.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.
KW Hypothetical protein.
SQ SEQUENCE 342 AA; 39470 MW; 85EFE2A77F8D766B CRC64;

Query Match 79.3%; Score 23; DB 5; Length 342;
Best Local Similarity 37.5%; Pred. No. 1.9e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 128 HSHAHTTH 135

RESULT 12
O54173 PRELIMINARY; PRT; 382 AA.
ID O54173
AC O54173;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein SCO5944.
GN SCO5944 OR SC7H1.14.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Thomson N.R., Brown S., Chandra G., Chen C.W., Collins M.,
RA Harper D., Bateman A., Brown S., Hidaigo J., Hornsby T., Howarth S.,
RA Cronin A., Fraser A., Goble A., Larke L., Murphy L., O'Neil S.,
RA Huang C.-H., Kieser T., Rajandream M.A., Rutherford K., Rutter S.,
RA Rabinowitsch E., Sharp S., Squares R., Taylor K.,
RA Seeger K., Saunders D., Woodward J., Barrell B.G., Parkhill J.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939125; CAA16201.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 382 AA; 41006 MW; 17C54D56069CE871 CRC64;

Query Match 79.3%; Score 23; DB 16; Length 382;
Best Local Similarity 37.5%; Pred. No. 2.1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 370 HAARHAAH 377

RESULT 13
Q9KX27 PRELIMINARY; PRT; 402 AA.
AC Q9KX27;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CoxC protein.
GN COXC.
OS Oligotropha carboxidovorans (Pseudomonas carboxydovorans).
OG Plasmid pHCG3.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Oligotropha.
OX NCBI_TaxID=40137;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OM5;
RA Moersdorf G.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OM5;
RA Schuebel U.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=OM5;
RX MEDLINE=95238294; PubMed=7721710;
RA Schuebel U., Kraut M., Moersdorf G., Meyer O.;
RT "Molecular Characterization of the Gene Cluster coxMSL Encoding the
RT Molybdenum-Containing Carbon Monoxide Dehydrogenase of Oligotropha
RT carboxidovorans.";
RL J. Bacteriol. 177:2197-2197(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=OM5;
RA Santiago B., Schuebel U., Egelseer C., Meyer O.;
RT "Sequence analysis, characterization and CO-specific transcription of

RT the cox gene cluster on the megaplasmid pHCG3 of Oligotropha
RT carboxidovorans.";
RL Gene 236:1157-1247(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=OM5;
RA Santiago B.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82447; CAB76246.1; -.
DR InterPro; IPR005330; SPNT_Repeat.
DR Pfam; PF03707; MHT; 4.
KW Plasmid.
SQ SEQUENCE 402 AA; 42864 MW; 4C6108F085CA133D CRC64;

Query Match 79.3%; Score 23; DB 2; Length 402;
Best Local Similarity 37.5%; Pred. No. 2.1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 127 HASAEHMT 134

RESULT 14
Q98UI5 PRELIMINARY; PRT; 412 AA.
AC Q98UI5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Thyroid transcription factor-1.
GN LJTTF-1.
OS Lampetra japonica (Japanese lamprey) (Entosphenus japonicus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Lethenteron.
OX NCBI_TaxID=94989;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ventral forebrain;
RA Ogasawara M., Shigetani Y., Suzuki S., Kuratani S., Satoh N.;
RT "Expression of Thyroid Transcription Factor-1 (TTF-1) Gene in the
RT Ventral Forebrain and Endostyle of the Agnathan Vertebrate, Lampetra
RT japonica.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AB052339; BAB32434.1; -.
DR HSSP; P23441; 1FTT.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 412 AA; 43509 MW; EC844185CD89D5EB CRC64;

Query Match 79.3%; Score 23; DB 13; Length 412;
Best Local Similarity 37.5%; Pred. No. 2.2e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 114 HHAHAAH 121

RESULT 15
Q947K4 PRELIMINARY; PRT; 466 AA.
ID Q947K4;
AC Q947K4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Thiohydroximate S-glucosyltransferase.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RA Marillia E.-F.A., MacPherson J.M., Tsang E.W.T., Van Audenhove K.,
RA Keller W.A., GrootWassink J.W.D.;
RT "Molecular cloning of a Brassica napus thiohydroximate S-
RT glucosyltransferase gene and its expression in Escherichia coli.";
RL Physiol. Plantarum 0:0-0(2001).
DR EMBL; AF304430; AAL09350.1; -.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 466 AA; 50826 MW; D5991B82129C2C1C CRC64;

Query Match 79.3%; Score 23; DB 10; Length 466;
Best Local Similarity 37.5%; Pred. No. 2.4e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
Db 188 HSSSHAHH 195

Search completed: November 12, 2003, 09:36:48
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:31:02 ; Search time 35.5 Seconds
(without alignments)
35.769 Million cell updates/sec

Title: US-10-064-903-2
Perfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	42	91.3	11	24	AAE29884	Clostridium tetani
2	42	91.3	439	23	ABG69073	Botulinum neurotox
3	42	91.3	441	23	ABG69068	Botulinum neurotox
4	42	91.3	441	23	ABG69076	Botulinum neurotox
5	42	91.3	441	23	ABB80654	Botulinum toxin ty
6	42	91.3	444	23	ABG69086	Botulinum neurotox
7	42	91.3	548	19	AAW56014	Recombinant botuli
8	42	91.3	848	23	ABG69087	Botulinum neurotox
9	42	91.3	852	23	ABG69077	Botulinum neurotox

10	42	91.3	858	19	AAW56018	Recombinant botuli
11	42	91.3	1169	19	AAW56017	Recombinant botuli
12	42	91.3	1291	19	AAW68392	Clostridium botuli
13	42	91.3	1315	22	AAB61169	Clostridium tetani
14	38	82.6	436	23	ABG69072	Botulinum neurotox
15	38	82.6	443	23	ABG69084	Botulinum neurotox
16	38	82.6	858	23	ABG69085	Botulinum neurotox
17	38	82.6	907	20	AAAY34888	Amino acid sequenc
18	36	78.3	104	23	ABP34326	Human ORF3299 prot
19	36	78.3	422	23	ABG69071	Botulinum neurotox
20	36	78.3	427	23	ABG69082	Botulinum neurotox
21	36	78.3	489	23	ABB55296	Lactococcus lactis
22	36	78.3	773	23	ABB91782	Herbicidally activ
23	36	78.3	804	23	ABG69083	Botulinum neurotox
24	35	76.1	60	15	AAR50716	G-protein coupled
25	35	76.1	60	17	AAW02908	G-protein coupled
26	35	76.1	155	23	ABU51691	Helicobacter pylor
27	35	76.1	240	23	ABU51965	Helicobacter pylor
28	35	76.1	251	22	ABB10271	Human cDNA SEQ ID
29	35	76.1	251	22	AAU22991	Novel human enzyme
30	35	76.1	251	23	ABP66858	Human polypeptide
31	35	76.1	269	14	AAR44805	Human cycilin D2 ps
32	35	76.1	269	24	ABU03642	Human expressed pr
33	35	76.1	315	18	AAW20813	H. pylori secreted
34	35	76.1	737	24	ABU00661	S. pneumoniae type
35	34	73.9	9	22	AAM98985	Vaccine related MH
36	34	73.9	118	22	AAB63297	Human breast cance
37	34	73.9	119	23	ABP34926	Human ORF3899 prot
38	34	73.9	150	22	AAB63290	Human breast cance
39	34	73.9	182	22	AAU37971	Streptococcus pneu
40	34	73.9	198	23	ABB47377	Listeria monocytog
41	34	73.9	198	24	ABU02143	S. pneumoniae type
42	34	73.9	226	19	AAV85978	S. pneumoniae deri
43	34	73.9	452	22	AAU38112	Salmonella typhi c
44	34	73.9	456	22	AAU34657	E. coli cellular p
45	34	73.9	498	22	ABG25738	Novel human diagno

ALIGNMENTS

RESULT 1
AAE29884
ID AAE29884 standard; peptide; 11 AA.
XX
AC AAE29884;
XX
DT 24-FEB-2003 (first entry)
XX
DE Clostridium tetani TeTx L chain fragment.

XX Gonadotrophin releasing hormone analogue; neurotoxin; prostate cancer;
KW endocrine disorder; gonadotrophin related illness; endometrial cancer;
KW pancreatic cancer; breast cancer; endometriosis; precocious puberty;
KW GnRH-A; therapy; protease; L chain; tetani toxin; Tefx.

XX Clostridium tetani.

XX WO200274327-A2.

PD 26-SEP-2002.

XX 11-MAR-2002; 2002WO-US07379.

XX 15-MAR-2001; 2001US-0810601.

XX (ALLR) ALLERGAN SALES INC.

PI Donovan S;

XX WPI; 2003-018772/01.

XX New agent comprising a light chain and a (modified) heavy chain of a

PT botulinum, butyricum, or tetani toxin, useful for treating a
PT gonadotrophin related illness, e.g. breast, prostate pancreatic or
PT endometrial cancer, or endometriosis -
XX
PS Disclosure; Fig 1B; 97pp; English.
XX
CC The invention relates to an agent comprising a neurotoxin preferably
CC botulinum toxin for treating endocrine disorders for e.g. gonadotrophin
CC related illness. The agent is useful for treating gonadotrophin related
CC illness e.g. prostate cancer, endometrial cancer, pancreatic cancer,
CC breast cancer, endometriosis or precocious puberty. It is also useful
CC for decreasing gonadotrophin secretion in a mammal. The present sequence
CC is Clostridium tetani Tefx (tetani toxin) L-chain fragment. This peptide
CC is used in the invention.
XX
SQ Sequence 11 AA;

Query Match 91.3%; Score 42; DB 24; Length 11;
Best Local Similarity 87.5%; Pred. No. 0.27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 4 HELIHVLH 11

RESULT 2
ABG69073
ID ABG69073 standard; Protein; 439 AA.
XX
AC ABG69073;

07-OCT-2002 (first entry)

Botulinum neurotoxin light chain polypeptide #7.

Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
lower motor neuron hyperactivity; autonomic nerve function; muscular;
immunostimulant; antibacterial.

Clostridium botulinum.

WO200236758-A2.

10-MAY-2002.

06-NOV-2001; 2001WO-US47230.

06-NOV-2000; 2000US-246774P.

20-JUL-2001; 2001US-0910186.

09-AUG-2001; 2001US-311966P.
(USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.

Smith LA, Jensen M;

WPI; 2002-575192/61.

N-PSDB; ABK98543.

Novel nucleic acid molecule encoding botulinum neurotoxin light chain
serotype A, useful for producing the neurotoxin for vaccination against
botulism, comprises sequence expressible in host other than Clostridium

Claim 33; Page 129-130; 166pp; English.

The invention relates to a nucleic acid molecule encoding a botulinum
neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
that is expressible in a host organism other than Clostridium, or has a
total A+T content that is less than about 70% The BoNT LC protein is

CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein.

SQ Sequence 439 AA;

Query Match 91.3%; Score 42; DB 23; Length 439;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 229 HELIHVLH 236

RESULT 3
ABG69068
ID ABG69068 standard; Protein; 441 AA.
XX
AC ABG69068;

07-OCT-2002 (first entry)

Botulinum neurotoxin light chain polypeptide #2.

Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
lower motor neuron hyperactivity; autonomic nerve function; muscular;
immunostimulant; antibacterial.

Clostridium botulinum.

WO200236758-A2.

10-MAY-2002.

06-NOV-2001; 2001WO-US47230.

06-NOV-2000; 2000US-246774P.

20-JUL-2001; 2001US-0910186.

09-AUG-2001; 2001US-311966P.
(USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.

Smith LA, Jensen M;

WPI; 2002-575192/61.

N-PSDB; ABK98538.

Novel nucleic acid molecule encoding botulinum neurotoxin light chain
serotype A, useful for producing the neurotoxin for vaccination against
botulism, comprises sequence expressible in host other than Clostridium

Claim 33; Page 119-120; 166pp; English.

The invention relates to a nucleic acid molecule encoding a botulinum
neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
that is expressible in a host organism other than Clostridium, or has a
total A+T content that is less than about 70% The BoNT LC protein is
useful in vaccination against botulism, for eliciting protective immunity
in a mammal, for treating dystonias, spasticity, pain, ocular motility,
facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
conditions characterised by hyperactivity of the lower motor neuron, and

CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein.
XX
SQ Sequence 441 AA;
Query Match 91.3%; Score 42; DB 23; Length 441;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237
RESULT 4
ABG69076
ID ABG69076 standard; Protein; 441 AA.
XX
AC ABG69076;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #10.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US47230.
XX
PR 06-NOV-2000; 2000US-246774P.
PR 20-JUL-2001; 2001US-0910186.
PR 09-AUG-2001; 2001US-311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98546.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium
PT -
XX
PS Claim 33; Page 135-136; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein.

SQ Sequence 441 AA;
Query Match 91.3%; Score 42; DB 23; Length 441;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237
RESULT 5
ABB80654
ID ABB80654 standard; peptide; 441 AA.
XX
AC ABB80654;
XX
DT 15-JUL-2002 (first entry)
XX
DE Botulinum toxin type B Danish I light chain.
XX
KW Neurotoxin; biological persistence; dysphonia; strabismus; muscle spasm;
KW dystonia; pain; blepharospasm; hemifacial spasm; excessive salivation;
KW eyelid disorder; cerebral palsy; focal spasticity; spasmodic colitis;
KW neurogenic bladder; anismus; limb spasticity; tic; tremor; bruxism;
KW anal fissure; achalasia; dysphagia; lacrimation; hyperhidrosis; headache;
KW excessive gastrointestinal secretion; botulinum toxin type B; Danish I;
KW light chain.
XX
OS Clostridium botulinum.
XX
PN WO200208268-A2.
XX
PD 31-JAN-2002.
XX
PF 20-JUL-2001; 2001WO-US23122.
XX
PR 21-JUL-2000; 2000US-0620840.
XX
PA (ALLR) ALLERGAN SALES INC.
XX
PI Steward LE, Fernandez-salas E, Herrington TM, Aoki KR;
XX
DR WPI; 2002-241566/29.
XX
PT Novel modified neurotoxin comprising structural modification which
PT alters the biological persistence and/or biological activity of a
PT neurotoxin, useful for treating neuromuscular or autonomic disorder, or
PT pain -
XX
PS Disclosure; Fig 8; 102pp; English.
XX
CC The sequence represents the botulinum toxin type B Danish I light chain.
CC The invention relates to a novel modified neurotoxin including a
CC structural modification, where the structural modification is effective
CC to alter the biological persistence, or biological activity. The modified
CC neurotoxin is useful for treating spasmodic dysphonia, laryngeal
CC dystonia, oromandibular dysphonia, lingual dystonia, cervical dystonia,
CC focal hand dystonia, blepharospasm, strabismus, hemifacial spasm, eyelid
CC disorder, cerebral palsy, focal spasticity, spasmodic colitis, neurogenic
CC bladder, anismus, limb spasticity, tics, tremors, bruxism, anal fissure,
CC achalasia, dysphagia, lacrimation, hyperhidrosis, excessive salivation,
CC excessive gastrointestinal secretions, pain from muscle spasms, headache
CC pain, brow furrows or skin wrinkles.
XX
SQ Sequence 441 AA;
Query Match 91.3%; Score 42; DB 23; Length 441;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237

Db 230 HELIHVLH 237

RESULT 6
ABG69086
ID ABG69086 standard; Protein; 444 AA.
XX
AC ABG69086;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #20.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US47230.
XX
PR 06-NOV-2000; 2000US-246774P.
PR 20-JUL-2001; 2001US-0910186.
PR 09-AUG-2001; 2001US-311966P.
XX
XX (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
PA Smith LA, Jensen M;
XX WPI; 2002-575192/61.
DR N-PSDB; ABK98556.
XX
XX Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium
PT -
XX
XX Claim 33; Page 160-161; 166pp; English.
PS
XX The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein.
XX
SQ Sequence 444 AA;
Query Match 91.3%; Score 42; DB 23; Length 444;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHVLH 237

RESULT 7
AAW56014

ID AAW56014 standard; Protein; 548 AA.
XX
AC AAW56014;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH107/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KW immunogen; detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-GB02273.
XX
PR 13-DEC-1996; 96GB-0025996.
PR 23-AUG-1996; 96GB-0017671.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
XX
PI Foster KA, Quinn CP, Shone CC;
XX
DR WPI; 1998-169168/15.
DR N-PSDB; AAV26286.
XX
PT Recombinant neurotoxin polypeptides - used to develop therapeutic
PT agents, immunogens or as non-toxic standards for the detection of
PT neurotoxins
XX
PS Example 2; Page 103-104; 137pp; English.
XX
CC The present sequence represents a recombinant neurotoxin protein from
CC the present invention. The present invention describes recombinant
CC neurotoxin proteins which comprise a first and second domain, where
CC the first domain is adapted to cleave one or more vesicle or
CC plasma-membrane associated proteins essential to exocytosis, and where
CC the second domain is adapted: (a) to translocate the protein into a
CC cell; (b) to increase the solubility of the protein compared to the
CC solubility of the first domain on its own, or (c) both to translocate
CC the protein into a cell and to increase the solubility of the protein
CC compared to the solubility of the first domain on its own, the protein
CC being free of clostridial neurotoxin (CN) and free of CN precursor that
CC can be converted into toxin by proteolytic action. The recombinant
CC proteins can be used as therapeutic agents for targeting cells
CC expressing a relevant substrate. The products can also be used as
CC immunogens and as non-toxic standards for the assessment and development
CC of in vitro assays for the detection of functional botulinum or tetanus
CC neurotoxins either in foodstuffs or in environmental samples.
XX
SQ Sequence 548 AA;
Query Match 91.3%; Score 42; DB 19; Length 548;
Best Local Similarity 87.5%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHVLH 237

RESULT 8
ABG69087
ID ABG69087 standard; Protein; 848 AA.
XX
AC ABG69087;
XX
DT 07-OCT-2002 (first entry)

XX Botulinum neurotoxin light chain polypeptide #21.
DE
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US47230.
XX
PR 06-NOV-2000; 2000US-246774P.
PR 20-JUL-2001; 2001US-0910186.
PR 09-AUG-2001; 2001US-311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98557.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium
PT -
XX
PS Claim 52; Page 162-164; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein.
XX
SQ Sequence 848 AA;

Query Match 91.3%; Score 42; DB 23; Length 848;
Best Local Similarity 87.5%; Pred. No. 25;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237

RESULT 9
ABG69077
ID ABG69077 standard; Protein; 852 AA.
XX
AC ABG69077;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #11.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;

KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US47230.
XX
PR 06-NOV-2000; 2000US-246774P.
PR 20-JUL-2001; 2001US-0910186.
PR 09-AUG-2001; 2001US-311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98547.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium
PT -
XX
PS Claim 52; Page 138-139; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein.
XX
SQ Sequence 852 AA;

Query Match 91.3%; Score 42; DB 23; Length 852;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 229 HELIHVLH 236

RESULT 10
AAW56018
ID AAW56018 standard; Protein; 858 AA.
XX
AC AAW56018;
XX
DT 27-JUL-1998 (first entry)
XX
DE Recombinant botulinum neurotoxin type B LH417/B.
XX
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KW immunogen; detection; tetanus; non-toxic; toxin.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO9807864-A1.

XX 26-FEB-1998.
XX PD
XX PF 22-AUG-1997; 97WO-GB02273.
XX PR 13-DEC-1996; 96GB-0025996.
XX PR 23-AUG-1996; 96GB-0017671.
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX (SPEY-) SPEYWOOD LAB LTD.
XX Foster KA, Quinn CP, Shone CC;
XX WPI; 1998-169168/15.
XX DR N-PSDB; AAV26290.
XX Recombinant neurotoxin polypeptides - used to develop therapeutic agents, immunogens or as non-toxic standards for the detection of neurotoxins
XX Example 2; Page 98-100; 137pp; English.
XX The present sequence represents a recombinant neurotoxin protein from the present invention. The present invention describes recombinant neurotoxin proteins which comprise a first and second domain, where the first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and where the second domain is adapted: (a) to translocate the protein into a cell; (b) to increase the solubility of the protein compared to the solubility of the first domain on its own, or (c) both to translocate the protein into a cell and to increase the solubility of the protein compared to the solubility of the first domain on its own, the protein being free of clostridial neurotoxin (CN) and free of CN precursor that can be converted into toxin by proteolytic action. The recombinant proteins can be used as therapeutic agents for targeting cells expressing a relevant substrate. The products can also be used as immunogens and as non-toxic standards for the assessment and development of in vitro assays for the detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples.
SQ Sequence 858 AA;
Query Match 91.3%; Score 42; DB 19; Length 858;
Best Local Similarity 87.5%; Pred. No. 26;
Matches .7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237
RESULT 11
AAW56017
ID AAW56017 standard; Protein; 1169 AA.
XX AC
XX AAW56017;
DT 27-JUL-1998 (first entry)
XX DE
XX Recombinant botulinum neurotoxin type B LH728/B.
XX Botulinum; recombinant; Clostridium botulinum; neurotoxin; immunogen; detection; tetanus; non-toxic; toxin.
XX Synthetic.
XX Clostridium botulinum.
XX WO9807864-A1.
XX PN
XX 26-FEB-1998.
XX PD
XX PF 22-AUG-1997; 97WO-GB02273.
XX

PR 13-DEC-1996; 96GB-0025996.
PR 23-AUG-1996; 96GB-0017671.
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX (SPEY-) SPEYWOOD LAB LTD.
XX Foster KA, Quinn CP, Shone CC;
XX WPI; 1998-169168/15.
XX DR N-PSDB; AAV26289.
XX Recombinant neurotoxin polypeptides - used to develop therapeutic agents, immunogens or as non-toxic standards for the detection of neurotoxins
XX Example 2; Page 91-94; 137pp; English.
XX The present sequence represents a recombinant neurotoxin protein from the present invention. The present invention describes recombinant neurotoxin proteins which comprise a first and second domain, where the first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and where the second domain is adapted: (a) to translocate the protein into a cell; (b) to increase the solubility of the protein compared to the solubility of the first domain on its own, or (c) both to translocate the protein into a cell and to increase the solubility of the protein compared to the solubility of the first domain on its own, the protein being free of clostridial neurotoxin (CN) and free of CN precursor that can be converted into toxin by proteolytic action. The recombinant proteins can be used as therapeutic agents for targeting cells expressing a relevant substrate. The products can also be used as immunogens and as non-toxic standards for the assessment and development of in vitro assays for the detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples.
SQ Sequence 1169 AA;
Query Match 91.3%; Score 42; DB 19; Length 1169;
Best Local Similarity 87.5%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237
RESULT 12
AAW68392
ID AAW68392 standard; Protein; 1291 AA.
XX AC
XX AAW68392;
XX DT 07-DEC-1998 (first entry)
XX DE
XX Clostridium botulinum type B toxin.
XX KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen; botulism; BotB.
XX OS Clostridium botulinum serotype B Danish strain.
XX PN WO9808540-A1.
XX PD 05-MAR-1998.
XX PF 28-AUG-1997; 97WO-US15394.
XX PR 28-AUG-1996; 96US-0704159.
XX (OPHI-) OPHIDIAN PHARM INC.
XX Thalley BS, Williams JA;
XX

SQ Sequence 436 AA;
Query Match 82.6%; Score 38; DB 23; Length 436;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
|:|||||
Db 225 HELIHALH 232

RESULT 15
ABG69084
ID ABG69084 standard; Protein; 443 AA.
XX
AC ABG69084;
XX
DT 07-OCT-2002 (first entry)
XX
DE Botulinum neurotoxin light chain polypeptide #18.
XX
KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;
KW immunostimulant; antibacterial.
XX
OS Clostridium botulinum.
XX
PN WO200236758-A2.
XX
PD 10-MAY-2002.
XX
PF 06-NOV-2001; 2001WO-US47230.
XX
PR 06-NOV-2000; 2000US-246774P.
PR 20-JUL-2001; 2001US-0910186.
PR 09-AUG-2001; 2001US-311966P.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Jensen M;
XX
DR WPI; 2002-575192/61.
DR N-PSDB; ABK98554.
XX
PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain
PT serotype A, useful for producing the neurotoxin for vaccination against
PT botulism, comprises sequence expressible in host other than Clostridium
PT
XX
PS Claim 33; Page 155-156; 166pp; English.
XX
CC The invention relates to a nucleic acid molecule encoding a botulinum
CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence
CC that is expressible in a host organism other than Clostridium, or has a
CC total A+T content that is less than about 70% The BoNT LC protein is
CC useful in vaccination against botulism, for eliciting protective immunity
CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,
CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental
CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,
CC conditions characterised by hyperactivity of the lower motor neuron, and
CC to control autonomic nerve function or tiptoe-walking due to stiff
CC muscles common in children with cerebral palsy. The sequences are also
CC useful for screening for botulinum neurotoxin inhibitors. This sequence
CC represents a botulinum neurotoxin light chain serotype A protein.
XX
SQ Sequence 443 AA;
Query Match 82.6%; Score 38; DB 23; Length 443;
Best Local Similarity 75.0%; Pred. No. 67;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 227 HELIHALH 234

Search completed: November 12, 2003, 09:35:07
Job time : 36.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: November 12, 2003, 09:33:48 ; Search time 14 Seconds
(without alignments)
24.178 Million cell updates/sec

Title: US-10-064-903-2
Perfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	548	4	US-09-255-829-24
2	42	91.3	858	4	US-09-255-829-22
3	42	91.3	858	4	US-09-255-829-29
4	42	91.3	1169	4	US-09-255-829-20
5	42	91.3	1315	4	US-08-913-880C-1
6	38	82.6	907	4	US-09-198-452A-306
7	35	76.1	60	1	US-08-118-270-257
8	35	76.1	60	5	PCT-US93-08528-257
9	35	76.1	269	5	PCT-US93-05000-31
10	34	73.9	204	4	US-09-107-532A-7103
11	33	71.7	60	1	US-08-117-083-20
12	33	71.7	65	1	US-07-879-685B-1
13	33	71.7	162	1	US-07-879-685B-4
14	33	71.7	416	1	US-08-117-083-62
15	33	71.7	431	1	US-08-311-023-2
16	33	71.7	541	2	US-08-540-804-16
17	33	71.7	541	2	US-08-218-265-16
18	33	71.7	541	3	US-08-521-872-16
19	33	71.7	541	3	US-08-590-399-16
20	33	71.7	942	4	US-09-657-931A-9
21	32	69.6	73	4	US-09-328-352-5131
22	32	69.6	285	4	US-09-328-352-7570
23	32	69.6	289	3	US-09-105-697-2
24	32	69.6	300	4	US-09-585-858-42
25	32	69.6	310	2	US-08-484-956-88
26	32	69.6	310	2	US-08-757-653-88
27	32	69.6	310	4	US-08-520-946-88

28	32	69.6	315	2	US-08-484-956-91	Sequence 91, Appl
29	32	69.6	315	2	US-08-757-653-91	Sequence 91, Appl
30	32	69.6	315	4	US-08-520-946-91	Sequence 91, Appl
31	32	69.6	320	2	US-08-757-653-163	Sequence 163, Appl
32	32	69.6	320	2	US-08-823-516-61	Sequence 61, Appl
33	32	69.6	320	3	US-08-759-038-102	Sequence 102, App
34	32	69.6	320	3	US-08-758-314-102	Sequence 102, App
35	32	69.6	320	4	US-09-684-938-102	Sequence 102, App
36	32	69.6	320	4	US-09-308-825A-102	Sequence 102, App
37	32	69.6	322	2	US-08-484-956-89	Sequence 89, Appl
38	32	69.6	322	2	US-08-757-653-89	Sequence 89, Appl
39	32	69.6	322	4	US-08-520-946-89	Sequence 89, Appl
40	32	69.6	359	4	US-09-252-991A-18134	Sequence 18134, A
41	32	69.6	528	2	US-08-484-956-90	Sequence 90, Appl
42	32	69.6	528	2	US-08-757-653-90	Sequence 90, Appl
43	32	69.6	528	4	US-08-520-946-90	Sequence 90, Appl
44	32	69.6	548	2	US-08-484-956-86	Sequence 86, Appl
45	32	69.6	548	2	US-08-757-653-86	Sequence 86, Appl

ALIGNMENTS

RESULT 1
US-09-255-829-24
; Sequence 24, Application US/092555829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-24

Query Match 91.3%; Score 42; DB 4; Length 548;
Best Local Similarity 87.5%; Pred. No. 8.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;


```
QY      1 HDLIHVLH 8
      |:|||||
Db      230 HELIHVLH 237

RESULT 2
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-22

Query Match      91.3%; Score 42; DB 4; Length 858;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
      |:|||||
Db      230 HELIHVLH 237

RESULT 3
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
```

```
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-29

Query Match      91.3%; Score 42; DB 4; Length 858;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
      |:|||||
Db      230 HELIHVLH 237

RESULT 4
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
```

ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1581.0130002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1169 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-255-829-20

Query Match 91.3%; Score 42; DB 4; Length 1169;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 5
US-08-913-880C-1
Sequence 1, Application US/08913880C
Patent No. 6372225
GENERAL INFORMATION:
APPLICANT: MATSUDA, Morihiro
TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
TITLE OF INVENTION: VACCINE
FILE REFERENCE: 216-380P
CURRENT APPLICATION NUMBER: US/08/913,880C
CURRENT FILING DATE: 1997-11-24
NUMBER OF SEQ ID NOS: 17
SEQ ID NO 1
LENGTH: 1315
TYPE: PRT
ORGANISM: Clostridium tetani
US-08-913-880C-1

Query Match 91.3%; Score 42; DB 4; Length 1315;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 233 HELIHLVH 240

RESULT 6
US-09-198-452A-306
Sequence 306, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffais, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
TITLE OF INVENTION: and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 306
LENGTH: 907
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-306

Query Match 82.6%; Score 38; DB 4; Length 907;
Best Local Similarity 62.5%; Pred. No. 65;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 541 HDLHLH 548

RESULT 7
US-08-118-270-257
Sequence 257, Application US/08118270
Patent No. 5508384
GENERAL INFORMATION:
APPLICANT: Murphy, Randall B.
APPLICANT: Schuster, David I.
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,270
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY=2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633

INFORMATION FOR SEQ ID NO: 257:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-118-270-257

Query Match 76.1%; Score 35; DB 1; Length 60;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 42 HDLVSILH 49

RESULT 8
PCT-US93-08528-257
Sequence 257, Application PC/TUS9308528
GENERAL INFORMATION:
APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.

;
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY=2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 257:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-257

Query Match 76.1%; Score 35; DB 5; Length 60;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|||: ||
Db 42 HDLVSLH 49

RESULT 9
PCT-US93-05000-31
; Sequence 31, Application PC/TUS9305000
; GENERAL INFORMATION:
; APPLICANT: MITOTIX
; TITLE OF INVENTION: D-Type Cyclin and Uses Related Thereto
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05000
; FILING DATE: 19930525
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,178
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL91-02A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 616-861-9540
; INFORMATION FOR SEQ ID NO: 31:

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US93-05000-31

Query Match 76.1%; Score 35; DB 5; Length 269;
Best Local Similarity 62.5%; Pred. No. 66;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|||: ||
Db 158 HDLIHVLH 165

RESULT 10
US-09-107-532A-7103
; Sequence 7103, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 7103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...204
; SEQUENCE DESCRIPTION: SEQ ID NO: 7103:
US-09-107-532A-7103

Query Match 73.9%; Score 34; DB 4; Length 204;
Best Local Similarity 75.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|||: |||

Db 107 HGLYHVLH 114

RESULT 11

US-08-117-083-20

; Sequence 20, Application US/08117083

; Patent No. 5719054

; GENERAL INFORMATION:

; APPLICANT: Boursnell, Michael E.

; APPLICANT: Inglis, Stephen C.

; APPLICANT: Munro, Alan J.

; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human

; TITLE OF INVENTION: Papilloma Virus Proteins

; NUMBER OF SEQUENCES: 70

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Walter H. Dreger

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/117,083

; FILING DATE: 10-SEP-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Walter H.

; REGISTRATION NUMBER: 24,190

; REFERENCE/DOCKET NUMBER: A-58783

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-781-1989

; TELEFAX: 415-398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 60 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FEATURE:

; NAME/KEY: Protein

; LOCATION: 1..60

; OTHER INFORMATION: /note= "Xaa refers to stop codon in

; OTHER INFORMATION: the open reading frame."

US-08-117-083-20

Query Match 71.7%; Score 33; DB 1; Length 60;

Best Local Similarity 85.7%; Pred. No. 34;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 DLIHVLH 8

Db 24 DLIHFLH 30

RESULT 12

US-07-879-685B-1

; Sequence 1, Application US/07879685B

; Patent No. 5296383

; GENERAL INFORMATION:

; APPLICANT: DAIKIN INDUSTRIES, LTD.

; TITLE OF INVENTION: A human centromere antigen

; TITLE OF INVENTION: polypeptide

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Umeda Center Building, 4-12

; STREET: Nakazaki-nishi, 2-chome

; CITY: Kita-ku

; STATE: Osaka

; COUNTRY: Japan

; ZIP: 530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/879,685B

; FILING DATE: 19920507

; CLASSIFICATION: 436

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 3-102517

; FILING DATE: 08-May-1991

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 65 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FRAGMENT TYPE: internal fragment

; ORIGINAL SOURCE:

; ORGANISM: human

US-07-879-685B-1

Query Match 71.7%; Score 33; DB 1; Length 65;

Best Local Similarity 83.3%; Pred. No. 37;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHV 6

Db 42 HDLVHV 47

RESULT 13

US-07-879-685B-4

; Sequence 4, Application US/07879685B

; Patent No. 5296383

; GENERAL INFORMATION:

; APPLICANT: DAIKIN INDUSTRIES, LTD.

; TITLE OF INVENTION: A human centromere antigen

; TITLE OF INVENTION: polypeptide

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Umeda Center Building, 4-12

; STREET: Nakazaki-nishi, 2-chome

; CITY: Kita-ku

; STATE: Osaka

; COUNTRY: Japan

; ZIP: 530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/879,685B

; FILING DATE: 19920507

; CLASSIFICATION: 436

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 3-102517

; FILING DATE: 08-May-1991

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 162 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-879-685B-4

Query Match 71.7%; Score 33; DB 1; Length 162;

Best Local Similarity 83.3%; Pred. No. 87;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHV 6
|||:|
Db 139 HDLVHV 144

RESULT 14

US-08-117-083-62
; Sequence 62, Application US/08117083
; Patent No. 5719054
; GENERAL INFORMATION:
; APPLICANT: Bourns, Michael E.
; APPLICANT: Inglis, Stephen C.
; APPLICANT: Munro, Alan J.
; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
; TITLE OF INVENTION: Papilloma Virus Proteins
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,083
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-58783
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 416 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..416
; OTHER INFORMATION: /note= "Xaa refers to stop codon in
; the open reading frame."
US-08-117-083-62

Query Match 71.7%; Score 33; DB 1; Length 416;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 DLIHVLH 8
|||||
Db 307 DLIHFLH 313

RESULT 15

US-08-311-023-2
; Sequence 2, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian

; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; TITLE OF INVENTION: BREAST TUMOURS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32,925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-311-023-2

Query Match 71.7%; Score 33; DB 1; Length 431;
Best Local Similarity 62.5%; Pred. No. 2.2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|||:|
Db 232 HDYHHILH 239

Search completed: November 12, 2003, 09:38:08
Job time : 14 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 12, 2003, 09:35:43 ; Search time 23 Seconds
(without alignments)
59.739 Million cell updates/sec

Title: US-10-064-903-2
Perfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	46	100.0	8	15	US-10-064-903-2
2	42	91.3	439	14	US-10-011-588-17
3	42	91.3	441	11	US-09-910-346C-20
4	42	91.3	441	14	US-10-011-588-7
5	42	91.3	441	14	US-10-011-588-23
6	42	91.3	444	14	US-10-011-588-43
7	42	91.3	458	12	US-10-241-596-114
8	42	91.3	548	12	US-10-241-596-24
9	42	91.3	848	14	US-10-011-588-45
10	42	91.3	852	14	US-10-011-588-25
11	42	91.3	858	12	US-10-241-596-22
12	42	91.3	860	12	US-10-241-596-175
13	42	91.3	862	12	US-10-241-596-94
14	42	91.3	862	12	US-10-241-596-171
15	42	91.3	862	12	US-10-241-596-173

16	42	91.3	864	12	US-10-241-596-102	Sequence 102, App
17	42	91.3	865	12	US-10-241-596-100	Sequence 100, App
18	42	91.3	866	12	US-10-241-596-88	Sequence 88, Appl
19	42	91.3	866	12	US-10-241-596-104	Sequence 104, App
20	42	91.3	867	12	US-10-241-596-80	Sequence 80, Appl
21	42	91.3	867	12	US-10-241-596-96	Sequence 96, Appl
22	42	91.3	867	12	US-10-241-596-98	Sequence 98, Appl
23	42	91.3	870	12	US-10-241-596-92	Sequence 92, Appl
24	42	91.3	871	12	US-10-241-596-84	Sequence 84, Appl
25	42	91.3	871	12	US-10-241-596-86	Sequence 86, Appl
26	42	91.3	871	12	US-10-241-596-90	Sequence 90, Appl
27	42	91.3	872	12	US-10-241-596-145	Sequence 145, App
28	42	91.3	876	12	US-10-241-596-82	Sequence 82, Appl
29	42	91.3	876	12	US-10-241-596-106	Sequence 106, App
30	42	91.3	876	12	US-10-241-596-108	Sequence 108, App
31	42	91.3	879	12	US-10-241-596-143	Sequence 143, App
32	42	91.3	887	12	US-10-241-596-147	Sequence 147, App
33	42	91.3	888	12	US-10-241-596-112	Sequence 112, App
34	42	91.3	1169	12	US-10-241-596-20	Sequence 20, Appl
35	42	91.3	1315	12	US-10-241-596-141	Sequence 141, App
36	42	91.3	1420	12	US-10-241-596-110	Sequence 110, App
37	38	82.6	436	14	US-10-011-588-15	Sequence 15, Appl
38	38	82.6	443	14	US-10-011-588-39	Sequence 39, Appl
39	38	82.6	858	14	US-10-011-588-41	Sequence 41, Appl
40	36	78.3	422	14	US-10-011-588-13	Sequence 13, Appl
41	36	78.3	427	14	US-10-011-588-35	Sequence 35, Appl
42	36	78.3	804	14	US-10-011-588-37	Sequence 37, Appl
43	35	76.1	251	9	US-09-764-853-579	Sequence 579, App
44	35	76.1	302	12	US-10-259-165-72	Sequence 72, Appl
45	35	76.1	302	12	US-10-259-165-410	Sequence 410, App

ALIGNMENTS

RESULT 1
US-10-064-903-2
; Sequence 2, Application US/10064903
; Publication No. US20030059912A1
; GENERAL INFORMATION:
; APPLICANT: Biotecon Gesellschaft fur biotechnologische Entwicklung und Consulting
; APPLICANT: mbH
; TITLE OF INVENTION: HYBRID PROTEIN FOR INHIBITING THE DEGRANULATION OF MASTOCYTES AN.
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: BIO-001PCT-CIP
; CURRENT APPLICATION NUMBER: US/10/064,903
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US 09/700,540
; PRIOR FILING DATE: 2001-01-19
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-10-064-903-2

Query Match 100.0%; Score 46; DB 15; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.8e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|||||||
Db 1 HDLIHVLH 8

RESULT 2
US-10-011-588-17
; Sequence 17, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody

```
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic botulinum neurotoxin light chain of
; OTHER INFORMATION: serotype G based on wild-type Clostridium
; OTHER INFORMATION: botulinum sequence
US-10-011-588-17

Query Match          91.3%; Score 42; DB 14; Length 439;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      229 HELIHLVH 236

RESULT 3
US-09-910-346C-20
; Sequence 20, Application US/09910346C
; Publication No. US20030027752A1
; GENERAL INFORMATION:
; APPLICANT: STEWARD, LANCE E
; APPLICANT: FERNANDEZ-SALAS, ESTER
; APPLICANT: HERRINGTON, TODD M
; APPLICANT: AOKI, KEI R
; TITLE OF INVENTION: Leucine-based motif and clostridial neurotoxins
; FILE REFERENCE: D-2885CIP
; CURRENT APPLICATION NUMBER: US/09/910,346C
; CURRENT FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/620,840
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-09-910-346C-20

Query Match          91.3%; Score 42; DB 11; Length 441;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      230 HELIHLVH 237

RESULT 4
US-10-011-588-7
; Sequence 7, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:22
US-10-011-588-23

Query Match          91.3%; Score 42; DB 14; Length 441;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      230 HELIHLVH 237

RESULT 5
US-10-011-588-23
; Sequence 23, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:22
US-10-011-588-23

Query Match          91.3%; Score 42; DB 14; Length 441;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      230 HELIHLVH 237
```

```
Db          230 HELIHVLH 237

RESULT 6
US-10-011-588-43
; Sequence 43, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:42
; NAME/KEY: UNSURE
; LOCATION: (442)...(443)
; OTHER INFORMATION: Any amino acid at each position
US-10-011-588-43

Query Match          91.3%; Score 42; DB 14; Length 444;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 HDLIHVLH 8
|:|||||
Db          230 HELIHVLH 237

RESULT 7
US-10-241-596-114
; Sequence 114, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 114
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-114

Query Match          91.3%; Score 42; DB 12; Length 548;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 HDLIHVLH 8
|:|||||
Db          230 HELIHVLH 237

RESULT 9
US-10-011-588-45
; Sequence 45, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
```

```
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-114

Query Match          91.3%; Score 42; DB 12; Length 458;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 HDLIHVLH 8
|:|||||
Db          232 HELIHVLH 239

RESULT 8
US-10-241-596-24
; Sequence 24, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-24

Query Match          91.3%; Score 42; DB 12; Length 548;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 HDLIHVLH 8
|:|||||
Db          230 HELIHVLH 237

RESULT 9
US-10-011-588-45
; Sequence 45, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
```



```

; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 848
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:44
US-10-011-588-45

Query Match          91.3%; Score 42; DB 14; Length 848;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
      |:|||||
Db      230 HELIHVLH 237

RESULT 10
US-10-011-588-25
; Sequence 25, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:24
US-10-011-588-25

Query Match          91.3%; Score 42; DB 14; Length 852;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
      |:|||||
Db      229 HELIHVLH 236

RESULT 11
US-10-241-596-22
; Sequence 22, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; APPLICANT: Microbiological Research Authority
```

```

; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-22

Query Match          91.3%; Score 42; DB 12; Length 858;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
      |:|||||
Db      230 HELIHVLH 237

RESULT 12
US-10-241-596-175
; Sequence 175, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 175
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-175

Query Match          91.3%; Score 42; DB 12; Length 860;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
      |:|||||
Db      230 HELIHVLH 237

RESULT 13
US-10-241-596-94
; Sequence 94, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
```

```
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 94
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-94
```

```
Query Match          91.3%; Score 42; DB 12; Length 862;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 HDLIHVLH 8
      |:|||||
Db      232 HELIHVLH 239
```

```
RESULT 14
US-10-241-596-171
; Sequence 171, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 171
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-171
```

```
Query Match          91.3%; Score 42; DB 12; Length 862;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 HDLIHVLH 8
      |:|||||
Db      230 HELIHVLH 237
```

```
RESULT 15
US-10-241-596-173
; Sequence 173, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 173
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-173
```

```
Query Match          91.3%; Score 42; DB 12; Length 862;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 HDLIHVLH 8
      |:|||||
Db      230 HELIHVLH 237
```

Search completed: November 12, 2003, 09:44:49
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:33:02 ; Search time 13.5 Seconds
(without alignments)
56.989 Million cell updates/sec

Title: US-10-064-903-2
Perfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	42	91.3	1268	2 S33411	botulinum neurotox
2	42	91.3	1291	1 A48940	bontoxilysin (EC 3
3	42	91.3	1291	2 I40631	non-protolytic bo
4	42	91.3	1297	2 S39791	neurotoxin - Clost
5	42	91.3	1315	1 BTCLTN	tentoxilysin (EC 3
6	38	82.6	925	2 A72096	ct234 hypothetical
7	38	82.6	925	2 E81573	conserved hypothet
8	38	82.6	925	2 E86527	CT234 hypothetical
9	38	82.6	1274	2 I40813	neurotoxin type F
10	36	78.3	232	2 C85585	unknown protein en
11	36	78.3	232	2 B90735	hypothetical prote
12	36	78.3	268	2 S73042	purine nucleoside
13	36	78.3	489	2 G86867	prophage ps3 prote
14	36	78.3	773	2 C84554	hypothetical prote
15	36	78.3	1251	2 JH0256	botulinum neurotox
16	36	78.3	1252	2 S21178	botulinum neurotox
17	35	76.1	312	2 C71806	hypothetical prote
18	35	76.1	312	2 G64712	toxR-activated gen
19	35	76.1	416	2 T45051	hypothetical prote
20	35	76.1	431	2 T18753	hypothetical prote
21	35	76.1	500	2 C75455	carboxypeptidase-r
22	35	76.1	679	2 H95036	glycosyl hydrolase
23	35	76.1	737	2 D97907	alpha-xylosidase (
24	34	73.9	198	2 F95194	recombination prot
25	34	73.9	198	2 C98061	recombination prot
26	34	73.9	198	2 AD1788	recombination prot
27	34	73.9	198	2 AE1412	recombination prot
28	34	73.9	199	2 JC5718	superoxide dismuta
29	34	73.9	421	2 C84555	hypothetical prote

30	34	73.9	446	2 B82282	exodeoxyribonuclea
31	34	73.9	449	2 AF0820	exodeoxyribonuclea
32	34	73.9	456	1 NCEC7	exodeoxyribonuclea
33	34	73.9	456	2 C91050	exonuclease VII la
34	34	73.9	456	2 H85894	exonuclease VII, 1
35	34	73.9	523	2 T04742	hypothetical prote
36	34	73.9	734	2 D95856	conserved hypothet
37	34	73.9	2241	2 S09811	hypothetical prote
38	33	71.7	29	2 C60110	repetitive protein
39	33	71.7	186	2 AF2083	hypothetical prote
40	33	71.7	229	2 T20722	hypothetical prote
41	33	71.7	297	2 E84237	hypothetical prote
42	33	71.7	352	2 T38311	protein kinase - f
43	33	71.7	455	2 H82881	cytosol aminopepti
44	33	71.7	459	2 AH0349	exodeoxyribonuclea
45	33	71.7	521	2 T27606	hypothetical prote

ALIGNMENTS

RESULT 1

S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228; EMD:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 91.3%; Score 42; DB 2; Length 1268;
Best Local Similarity 87.5%; Pred. No. 6.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8

Db 219 HELIHLVH 226

RESULT 2

A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the ty
A;Reference number: A48940; MUID:92384550; PMID:1514783
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparison
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217, 'G', 219-224, 'S', 226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K.
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus toxin
A;Reference number: A42871; MUID:92340509; PMID:1634516
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313, 'S', 315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987; PMID:3139097
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29, 'M', 31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463, 'R', 465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963; PMID:3888113
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synapses
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVV>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 91.3%; Score 42; DB 1; Length 1291;
Best Local Similarity 87.5%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 3
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulin
A;Reference number: I40631; MUID:94122659; PMID:7764370
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulin neurotoxin gene and specific i
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761, 'E', 763-841, 'M', 843, 'T', 845, 'N', 847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843, 'T', 845, 'N', 847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may possess a capable catalytic
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVV>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 4
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium
A;Reference number: S39791; MUID:94092745; PMID:8268233
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 91.3%; Score 42; DB 2; Length 1297;
Best Local Similarity 87.5%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHVLH 237

RESULT 5
BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text_change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
A;Experimental source: strain CN3911
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 165, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schiaivo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation
R;de Filippis, V.; Vangelista, L.; Schiaivo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)
C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio
C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized
presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy
C;Function:
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt
C;Superfamily: tetanus toxin

C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylysin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylysin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 91.3%; Score 42; DB 1; Length 1315;
Best Local Similarity 87.5%; Pred. No. 6.4;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 233 HELIHVLH 240

RESULT 6

A72096
ct234 hypothetical protein - Chlamydophila pneumoniae (strain CWL029)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C;Accession: A72096
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606; PMID:10192388
A;Accession: A72096
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-925 <ARN>
A;Cross-references: GB:AE001614; GB:AE001363; NID:g4376562; PIDN:AAD18442.1; PID:g437656
A;Experimental source: strain CWL029
C;Genetics:
A;Gene: Cpn0293
C;Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 82.6%; Score 38; DB 2; Length 925;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:
Db 538 HDLKHITH 545

RESULT 7

E81573
conserved hypothetical protein CP0465 [imported] - Chlamydophila pneumoniae (strain AR3)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 20-Jun-2000
C;Accession: E81573
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: E81573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-925 <REA>
A;Cross-references: GB:AE002208; GB:AE002161; NID:g7189387; PIDN:AAF38302.1; PID:g718938
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: CP0465
C;Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 82.6%; Score 38; DB 2; Length 925;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:
Db 538 HDLKHITH 545

Db 538 HDLLHITH 545

RESULT 8
E86527
CT234 hypothetical protein [imported] - Chlamydomophila pneumoniae (strain J138)
C/Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C/Accession: E86527
R/Shirai, M.; Harakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A/Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A/Reference number: A86491; MUID:20330349; PMID:10871362
A/Accession: E86527
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-925 <STO>
A/Cross-references: GB:BA000008; NID:g8978667; PIDN:BAA98503.1; GSPDB:GN00142
A/Experimental source: strain J138
C/Genetics:
A/Gene: CPj0293
C/Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 82.6%; Score 38; DB 2; Length 925;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 538 HDLLHITH 545

RESULT 9
I40813
neurotoxin type F - Clostridium botulinum
C/Species: Clostridium botulinum
C/Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C/Accession: I40813; S48108
R/East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A/Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A/Reference number: I40644
A/Accession: I40813
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-1274 <RES>
A/Cross-references: GB:M92906; NID:gl44866; PIDN:AAA23263.1; PID:gl44867
R/Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A/Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A/Reference number: S48103; MUID:94013372; PMID:8408542
A/Accession: S48108
A/Status: preliminary; translation not shown
A/Molecule type: DNA
A/Residues: 634-1002 <CAM>
A/Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C/Superfamily: tetanus toxin
C/Keywords: neurotoxin

Query Match 82.6%; Score 38; DB 2; Length 1274;
Best Local Similarity 75.0%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 227 HELIHALH 234

RESULT 10
C85585
unknown protein encoded by prophage CP-933K [imported] - Escherichia coli (strain O157:H
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: C85585
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: C85585
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-232 <STO>
A/Cross-references: GB:AE005174; NID:g12513758; PIDN:AAG55143.1; GSPDB:GN00145; UWGP:Z0
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: Z0990

Query Match 78.3%; Score 36; DB 2; Length 232;
Best Local Similarity 62.5%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 142 HELLHVFFH 149

RESULT 11
B90735
hypothetical protein ECs0850 [imported] - Escherichia coli (strain O157:H7, substrain F
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C/Accession: B90735
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ger
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: B90735
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-232 <HAY>
A/Cross-references: GB:BA000007; PIDN:BAB34273.1; PID:gl3360309; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain RIMD 0509952
C/Genetics:
A/Gene: ECs0850

Query Match 78.3%; Score 36; DB 2; Length 232;
Best Local Similarity 62.5%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 142 HELLHVFFH 149

RESULT 12
S73042
purine nucleoside phosphorylase pnpH - Mycobacterium leprae
N/Alternate names: L308_F2_56 protein
C/Species: Mycobacterium leprae
C/Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C/Accession: S73042
R/Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A/Description: Mycobacterium leprae cosmid L308.
A/Reference number: S72590
A/Accession: S73042
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-268 <SMI>
A/Cross-references: EMBL:U00022; NID:g467164; PIDN:AAA17341.1; PID:g467183
C/Genetics:
A/Gene: pnpH
A/Start codon: GTG
C/Superfamily: purine-nucleoside phosphorylase

Query Match 78.3%; Score 36; DB 2; Length 268;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
||| |||
Db 95 HDLRHVH 102

RESULT 13
G86867
prophage ps3 protein 11 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: G86867
R;Bolotin, A.; Wincker, P.; Mauder, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: G86867
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-489 <STO>
A;Cross-references: GB:AE005176; PID:g12724983; PIDN:AAK06041.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ps311

Query Match 78.3%; Score 36; DB 2; Length 489;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 DLIHVLH 8
||:|:|
Db 49 DLVHILH 55

RESULT 14
C84554
hypothetical protein At2g17610 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84554
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-773 <STO>
A;Cross-references: GB:AE002093; NID:g4926870; PIDN:AAD32950.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g17610
A;Map position: 2

Query Match 78.3%; Score 36; DB 2; Length 773;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
||:|:|
Db 10 HELIHSLSH 17

RESULT 15
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999

C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428; PMID:1543481
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E tox
A;Reference number: S16145; MUID:91237316; PMID:2033376
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 78.3%; Score 36; DB 2; Length 1251;
Best Local Similarity 75.0%; Pred. No. 85;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
||:|:|
Db 212 HELIHSLSH 219

Search completed: November 12, 2003, 09:37:29
Job time : 14.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:31:57 ; Search time 10 Seconds
(without alignments)
37.621 Million cell updates/sec

Title: US-10-064-903-2
Perfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	1290	1 BXB_CLOBO	P10844 clostridium
2	42	91.3	1296	1 BXG_CLOBO	Q60393 clostridium
3	42	91.3	1314	1 TETX_CLOTE	P04958 clostridium
4	38	82.6	1274	1 BXF_CLOBO	P30996 clostridium
5	36	78.3	268	1 PUNA_MYCLE	P46862 mycobacteri
6	36	78.3	1250	1 BXE_CLOBO	Q00496 clostridium
7	36	78.3	1250	1 BXE_CLOBU	P30995 clostridium
8	34	73.9	198	1 RECR_LISIN	Q927d9 listeria in
9	34	73.9	198	1 RECR_LISMO	Q8y3x7 listeria mo
10	34	73.9	198	1 RECR_STRPN	Q9zhc4 streptococc
11	34	73.9	199	1 SODC_HAEDU	Q59452 haemophilus
12	34	73.9	446	1 EX7L_VIBCH	Q9ktw4 vibrio chol
13	34	73.9	449	1 EX7L_SALTI	Q8z4q1 salmonella
14	34	73.9	449	1 EX7L_SALTY	Q8zn58 salmonella
15	34	73.9	456	1 EX7L_ECO57	Q8xab0 escherichia
16	34	73.9	456	1 EX7L_ECOLI	P04994 escherichia
17	34	73.9	458	1 EX7L_ECOLI6	Q8ff64 escherichia
18	34	73.9	2241	1 TEGU_HCMVA	P16785 human cytom
19	33	71.7	239	1 CENB_SHEEP	P49451 ovis aries
20	33	71.7	352	1 PRK1_SCHPO	O13958 schizosacch
21	33	71.7	459	1 EX7L_YERPE	Q8zcu2 yersinia pe
22	33	71.7	541	1 UME5_YEAST	P39073 saccharomyc
23	33	71.7	599	1 CENB_HUMAN	P07199 homo sapien
24	33	71.7	599	1 CENB_MOUSE	P27790 mus musculu
25	33	71.7	606	1 CENB_CRIGR	P48988 cricetulus
26	33	71.7	697	1 AD26_MOUSE	Q9r158 mus musculu
27	33	71.7	740	1 FAS_PNECA	P29251 p folic aci
28	33	71.7	942	1 AMPN_MANSE	P91885 manduca sex
29	32	69.6	91	1 YVBG_VACCC	P20547 vaccinia vi
30	32	69.6	239	1 MT04_ECOLI	P32049 escherichia
31	32	69.6	477	1 GLGA_CLOAB	Q97gx6 clostridium
32	32	69.6	528	1 UGA3_YEAST	P26370 saccharomyc
33	32	69.6	612	1 OCTC_RAT	P11466 rattus norv

RESULT 1					ALIGNMENTS				
BXB_CLOBO					STANDARD; PRT; 1290 AA.				
ID	BXB_CLOBO	STANDARD;	PRT;	1290 AA.					
AC	P10844; P10843;								P19821 thermus aqu
DT	01-JUL-1989 (Rel. 11, Created)								P26607 escherichia
DT	01-JUL-1993 (Rel. 26, Last sequence update)								P59342 shigella fl
DT	15-SEP-2003 (Rel. 42, Last annotation update)								P53550 saccharomyc
DE	Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)								P21783 xenopus lae
DE	(Bontoxilysin B).								O66407 aquifex aeo
GN	BOTB.								Q9vvg6 drosophila
OS	Clostridium botulinum.								Q8y3a6 ralstonia s
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;								Q9rkn3 streptomyce
OC	Clostridium.								Q98et0 rhizobium l
OX	NCBI_TaxID=1491;								Q9hiw5 thermoplasm
RN	[1]								Q9plm9 chlamydia m
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=92384550; PubMed=1514783;								
RA	Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,								
RA	Minton N.P.;								
RT	"Molecular cloning of the Clostridium botulinum structural gene								
RT	encoding the type B neurotoxin and determination of its entire								
RT	nucleotide sequence.";								
RL	Appl. Environ. Microbiol. 58:2345-2354(1992).								
RN	[2]								
RP	SEQUENCE OF 35-245 FROM N.A.								
RC	STRAIN=NCTC 7273;								
RA	Szabo E.A., Pemberton J.M., Desmarchelier P.M.;								
RL	Submitted (APR-1992) to the EMBL/GenBank/DBSJ databases.								
RN	[3]								
RP	SEQUENCE OF 633-993 FROM N.A.								
RC	STRAIN=NCTC 7273;								
RX	MEDLINE=94013372; PubMed=8408542;								
RA	Campbell K., East A.K., Collins M.D.;								
RT	"Gene probes for identification of the botulinal neurotoxin gene and								
RT	specific identification of neurotoxin types B, E, and F.";								
RL	J. Clin. Microbiol. 31:2255-2262(1993).								
RN	[4]								
RP	SEQUENCE OF 1-44 AND 441-466.								
RC	STRAIN=657;								
RX	MEDLINE=89000987; Pubmed=3139097;								
RA	Dasgupta B.R., Datta A.;								
RT	"Botulinum neurotoxin type B (strain 657): partial sequence and								
RT	similarity with tetanus toxin.";								
RL	Biochimie 70:811-817(1988).								
RN	[5]								
RP	SEQUENCE OF 1-16 AND 441-458.								
RC	STRAIN=OKRA;								
RX	MEDLINE=85197963; PubMed=3888113;								
RA	Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;								
RT	"Partial amino acid sequences of botulinum neurotoxins types B and								
RT	E.";								
RL	Arch. Biochem. Biophys. 238:544-548(1985).								
RN	[6]								
RP	IDENTIFICATION AS ZINC-PROTEASE.								
RX	MEDLINE=93054694; Pubmed=1429690;								
RA	Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;								

FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 230 230 BY SIMILARITY.
FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 435 449 INTERCHAIN (PROBABLE).
SQ SEQUENCE 1296 AA; 149013 MW; DC8E47B15F665C31 CRC64;

Query Match 91.3%; Score 42; DB 1; Length 1296;
Best local similarity 87.5%; Pred. No. 2.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 229 HELIHVLH 236

RESULT 3
TETX CLOTE STANDARD; PRT; 1314 AA.
AC P04958;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylysin) [Contains:
DE Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy
DE chain (Tetanus toxin chain H)].
GN TETX OR CTP60.
OS Clostridium tetani.
OG Plasmid pE88, and Plasmid 75 Kbp.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC PLASMID=75 Kbp;
RX MEDLINE=87053814; PubMed=3536478;
RA Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
RA Weller U., Hudel M., Habermann E., Niemann H.;
RT "Tetanus toxin: primary structure, expression in E. coli, and
RT homology with botulinum toxins.";
RL EMBO J. 5:2495-2502(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CN3911; PLASMID=75 Kbp;
RX MEDLINE=87040747; PubMed=3774547;
RA Fairweather N.F., Lyness V.A.;
RT "The complete nucleotide sequence of tetanus toxin.";
RL Nucleic Acids Res. 14:7809-7812(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88; PLASMID=pE88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baeumer S., Fricke W.F., Wierzer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
RN [4]
RP SEQUENCE OF 742-1314 FROM N.A.
RC PLASMID=75 Kbp;
RX MEDLINE=86085672; PubMed=3510187;
RA Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;
RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT fragment C in Escherichia coli.";
RL J. Bacteriol. 165:21-27(1986).
RN [5]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=90201034; PubMed=2108021;
RA Krieglststein K., Henschen A., Weller U., Habermann E.;
RT "Arrangement of disulfide bridges and positions of sulphydryl groups
RT in tetanus toxin.";
RL Eur. J. Biochem. 188:39-45(1990).
RN [6]

RP PARTIAL SEQUENCE.
RX MEDLINE=92037649; PubMed=1935979;
RA Krieglststein K.G., Henschen A.H., Weller U., Habermann E.;
RT "Limited proteolysis of tetanus toxin. Relation to activity and
RT identification of cleavage sites.";
RL Eur. J. Biochem. 202:41-51(1991).
RN [7]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93010948; PubMed=1396558;
RA Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,
RA Montecucco C.;
RT "Tetanus toxin is a zinc protein and its inhibition of
RT neurotransmitter release and protease activity depend on zinc.";
RL EMBO J. 11:3577-3583(1992).
RN [8]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin.";
RL Nature 359:832-835(1992).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX MEDLINE=97475217; PubMed=9334741;
RA Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA Sax M.;
RT "Structure of the receptor binding fragment HC of tetanus
RT neurotoxin.";
RL Nat. Struct. Biol. 4:788-792(1997).
CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC BOND OF SYNAPTOSREVIN-2.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in
CC synaptobrevin 2.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC AND ARE NON-TOXIC AFTER SEPARATION.
CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC GLYNGLIOSIDE RECEPTORS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04436; CAA28033.1; -.
DR EMBL; X06214; CAA29564.1; -.
DR EMBL; AF528037; AAO37454.1; -.
DR EMBL; M12739; AAA23282.1; -.
DR PIR; A25689; BTCLTN.
DR PDB; 1AF9; 29-APR-98.
DR PDB; 1A8D; 14-OCT-98.
DR PDB; 1D0H; 27-MAR-00.
DR PDB; 1DFQ; 24-MAR-00.
DR PDB; 1DIW; 24-MAR-00.
DR PDB; 1DLL; 24-MAR-00.
DR PDB; 1FV3; 05-SEP-01.
DR MEROPS; M27.001; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD0019e3; Bontoxilysin; 1.
DR


```
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M92906; AAA23263.1; -.
CC EMBL; S73676; AAC60475.1; -.
CC EMBL; X70820; CAA50151.1; -.
CC EMBL; X70816; CAA50147.1; -.
CC PIR; I40813; I40813.
CC PIR; S48109; S48109.
CC HSSP; P10845; 3BTA.
CC MEROPS; M27.002; -.
CC InterPro; IPR000395; Bontoxilysin.
CC InterPro; IPR006025; Zn_MTpeptidse.
CC Pfam; PF01742; Peptidase_M27; 1.
CC ProDom; PD001963; Bontoxilysin; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT CHAIN 437 1274 BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT METAL 227 227 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 228 228 BY SIMILARITY.
FT METAL 231 231 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 429 445 INTERCHAIN (PROBABLE).
FT SEQUENCE 1274 AA; 146709 MW; 5B99756A7438B921 CRC64;

Query Match 82.6%; Score 38; DB 1; Length 1274;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 227 HELIHALH 234

RESULT 5
PUNA MYCLE
ID PUNA MYCLE STANDARD; PRT; 268 AA.
AC P46862;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase)
DE (PNP).
GN PUNA OR DEOD OR ML0707 OR L308_F2_56.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
```

```
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: CLEAVAGE OF GUANOSINE OR INOSINE TO RESPECTIVE BASES AND
CC SUGAR-1-PHOSPHATE MOLECULES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Purine nucleoside + phosphate = purine +
CC alpha-D-ribose 1-phosphate.
CC -!- PATHWAY: Purine nucleoside salvage.
CC -!- SIMILARITY: BELONGS TO THE PNP/MTAP FAMILY 2 OF PHOSPHORYLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00022; AAA17341.1; -.
CC EMBL; AL583919; CAC30216.1; -.
CC PIR; S73042; S73042.
CC HSSP; P81989; IQE5.
CC Leproma; ML0707; -.
CC InterPro; IPR001369; Mtap_PNP.
CC Pfam; PF00896; Mtap_PNP; 1.
CC PROSITE; PS01240; PNP_MTAP_2; 1.
KW Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 268 AA; 27980 MW; 46C622532FC96A0F CRC64;

Query Match 78.3%; Score 36; DB 1; Length 268;
Best Local Similarity 75.0%; Pred. No. 7.2;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 95 HDLRHVVH 102

RESULT 6
BXE CLOBO
ID BXE CLOBO STANDARD; PRT; 1250 AA.
AC Q00496;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin E).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beluga;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulin neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92174922; PubMed=1541280;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type-E
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 204:657-667(1992).
RN [3]
RP SEQUENCE OF 1-251 FROM N.A.
RX MEDLINE=90264400; PubMed=2160960;
RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
```


RN RP
RX SEQUENCE OF 1-13.
RA MEDLINE=85197963; PubMed=3888113;
Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
"Partial amino acid sequences of botulinum neurotoxins types B and E.";
Arch. Biochem. Biophys. 238:544-548(1985).
[5]
RN RP
RX SEQUENCE OF 419-426.
RA MEDLINE=90344918; PubMed=2116911;
Gimenez J.A., Dasgupta B.R.;
"Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the site trypsin nicks and homology with tetanus neurotoxin.";
Biochimie 72:213-217(1990).
[6]
RN RP
RX IDENTIFICATION OF SUBSTRATE.
MEDLINE=94063091; PubMed=8243676;
Schiaivo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J., Benfenati F., Wilson M.C., Montecucco C.;
"Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct COOH-terminal peptide bonds.";
FEBS Lett. 335:99-103(1993).
[7]
RN RP
RX IDENTIFICATION OF SUBSTRATE.
MEDLINE=94124495; PubMed=8294407;
Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C., Jahn R., Niemann H.;
"Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
J. Biol. Chem. 269:1617-1620(1994).
-!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-181 BOND IN SNAP-25.
-!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.
-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
-!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
-!- SUBCELLULAR LOCATION: Secreted.
-!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; X62089; CAA43999.1; --
DR EMBL; X62683; CAA44558.1; --
DR PIR; S21178; S21178.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; --
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTPeptidse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT MET 0 0
FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT CHAIN.
FT CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY CHAIN.

FT METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 212 212 BY SIMILARITY.
FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 411 425 INTERCHAIN (PROBABLE).
FT CONFLICT 176 176 R -> G (IN REF. 2).
FT CONFLICT 197 197 C -> S (IN REF. 2 AND 3).
FT CONFLICT 339 339 R -> A (IN REF. 2).
FT CONFLICT 772 772 I -> L (IN REF. 2).
FT CONFLICT 962 963 FE -> LQ (IN REF. 2).
FT CONFLICT 966 966 R -> A (IN REF. 2).
FT CONFLICT 1194 1194 N -> NN (IN REF. 2).
SQ SEQUENCE 1250 AA; 143712 MW; D9FCE26DDA041EB4 CRC64;
Query Match 78.3%; Score 36; DB 1; Length 1250;
Best Local Similarity 75.0%; Pred.No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 211 HELIHSLSH 218
RESULT 7
BXE_CLOBU STANDARD; PRT; 1250 AA.
AC P30395;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin E).
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43181, and ATCC 43755;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
"Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type E (strain Beluga) and Clostridium butyricum (strains ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE OF 1-251 FROM N.A.
RC STRAIN=BL6340;
RX MEDLINE=91237316; PubMed=2033376;
RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K., Yokosawa N., Yashiki T., Oguma K.;
"Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxin gene from Clostridium butyricum strain BL6340.";
RL J. Gen. Microbiol. 137:519-525(1991).
RN [3]
RP SEQUENCE OF 1-48.
RC STRAIN=5262;
RA Gimenez J., Foley J., Dasgupta B.R.;
"Neurotoxin type E from Clostridium botulinum and C. butyricum; partial sequence and comparison.";
RL FASEB J. 2:A1750-A1750(1988).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC ENDOPEPTIDASE.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL

CC FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X62088; CAA43998.1; -.

CC EMBL; X53180; CAA37321.1; -.

CC PIR; JH0256; JH0256.

CC HSSP; P10845; 3BTA.

CC MEROPS; M27.002; -.

CC InterPro; IPR000395; Bontoxilysin.

CC InterPro; IPR006025; Zn_MTpeptdse.

CC Pfam; PF01742; Peptidase M27; 1.

CC PRINTS; PR00760; BONTOXILYSIN.

CC ProDom; PD001963; Bontoxilysin; 1.

CC PROSITE; PS00142; ZINC_PROTEASE; 1.

CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.

KW INIT MET 0

FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.

FT CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.

FT METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 212 212 BY SIMILARITY.

FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).

FT DISULFID 411 425 INTERCHAIN (PROBABLE).

FT CONFLICT 229 229 K -> M (IN REF. 2).

FT SEQUENCE 1250 AA; 143265 MW; 8171B5B2C2312857 CRC64;

CC -----

CC Query Match 78.3%; Score 36; DB 1; Length 1250;

CC Best Local Similarity 75.0%; Pred. No. 37;

CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8

Db 211 HELIHSLSH 218

RESULT 8

RECR LISIN STANDARD; PRT; 198 AA.

ID RECR LISIN STANDARD; PRT; 198 AA.

AC Q927D9;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Recombination protein recr.

GN RECR OR LIN2850.

OS Listeria innocua.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI_TaxID=1642;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLIP 11262 / Serovar 6a;

RX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,

RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,

RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,

RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,

RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,

RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,

RT "Comparative genomics of Listeria species.";

RL Science 294:849-852(2001).

CC -!- FUNCTION: May play a role in DNA repair. It seems to be involved

CC in an recBC-independent recombinational process of DNA repair. It

CC may act with recF and reco (By similarity).

CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AL596173; CAC98076.1; -.

CC PIR; AD1788; AD1788.

CC ListList; LIN02850; -.

CC HAMAP; MF_00017; -; 1.

CC InterPro; IPR003583; HHH 1.

CC InterPro; IPR000093; RecR.

CC InterPro; IPR006171; Toprim_dom.

CC InterPro; IPR006154; Toprim_sub.

CC Pfam; PF02132; RecR; 1.

CC Pfam; PF01751; Toprim; 1.

CC SMART; SM00278; HhH1; 1.

CC SMART; SM00493; TOPRIM; 1.

CC TIGRFAMS; TIGR00615; recR; 1.

CC PROSITE; PS01300; RECR; 1.

CC DNA repair; DNA recombination; Zinc-finger; Complete proteome.

KW ZN FING 57 72 C4-TYPE (POTENTIAL).

FT SEQUENCE 198 AA; 21996 MW; 8A82E1A16415DFEF CRC64;

CC -----

CC Query Match 73.9%; Score 34; DB 1; Length 198;

CC Best Local Similarity 75.0%; Pred. No. 12;

CC Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8

Db 101 HGLYHVLH 108

RESULT 9

RECR LISMO STANDARD; PRT; 198 AA.

ID RECR LISMO STANDARD; PRT; 198 AA.

AC Q8Y3X7;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Recombination protein recr.

GN RECR OR LMO2702.

OS Listeria monocytogenes.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI_TaxID=1639;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EGD-e / Serovar 1/2a;

RX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,

RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,

RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,

RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,

RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,

RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,

RT "Comparative genomics of Listeria species.";

CC -!- FUNCTION: May play a role in DNA repair. It seems to be involved

CC in an recBC-independent recombinational process of DNA repair. It

CC may act with recF and reco (By similarity).

CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL591984; CAD00915.1; -.
DR PIR; AE1412; AE1412.
DR List; LMO02702; -.
DR HAMAP; MF 00017; -; 1.
DR InterPro; IPR003583; HHH 1.
DR InterPro; IPR000093; RecR.
DR InterPro; IPR006171; Toprim_dom.
DR InterPro; IPR006154; Toprim_sub.
DR Pfam; PF02132; RecR; 1.
DR Pfam; PF01751; Toprim; 1.
DR SMART; SM00278; HH1; 1.
DR SMART; SM00493; TOPRIM; 1.
DR TIGRFAMS; TIGR00615; RecR; 1.
DR PROSITE; PS01300; RecR; 1.
DR DNA repair; DNA recombination; Zinc-finger; Complete proteome.
KW C4-TYPE (POTENTIAL).
FT ZN FING 57 72
SQ SEQUENCE 198 AA; 21934 MW; E542E27BC3D05036 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 198;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
| | | | |
Db 101 HGLYHVLH 108

RESULT 10
RECR_STRPN STANDARD; PRT; 198 AA.
AC Q9ZHC4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Recombination protein recr.
GN RECR OR RECM OR SP1672 OR SPR1516.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G54 / Type 19F;
RX MEDLINE=99061199; PubMed=9846742;
RA Massidda O., Anderluzzi D., Friedli L., Feger G.;
RT "Unconventional organization of the division and cell wall gene
cluster of Streptococcus pneumoniae.";
RL Microbiology 144:3069-3078(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=339, and PN94-661;
RX MEDLINE=20073037; PubMed=10605111;
RA Enright M.C., Spratt B.G.;
RT "Extensive variation in the ddl gene of penicillin-resistant
Streptococcus pneumoniae results from a hitchhiking effect driven by
the penicillin-binding protein 2b gene.";
RL Mol. Biol. Evol. 16:1687-1695(1999).
RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
PRT; 199 AA.

RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae.";
RL Science 293:498-506(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
CC -!- FUNCTION: MAY PLAY A ROLE IN DNA REPAIR. IT SEEMS TO BE INVOLVED
CC IN AN RECB-INDEPENDENT RECOMBINATIONAL PROCESS OF DNA REPAIR. IT
CC MAY ACT WITH RECF AND RECO (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF068901; AAC95434.1; -.
DR EMBL; AJ243056; CAB64474.1; -.
DR EMBL; AJ243057; CAB64478.1; -.
DR EMBL; AE007460; AAK75751.1; -.
DR EMBL; AE008520; AAL00320.1; -.
DR PIR; C98061; C98061.
DR PIR; F95194; F95194.
DR TIGR; SP1672; -.
DR HAMAP; MF_00017; -; 1.
DR InterPro; IPR000093; RecR.
DR InterPro; IPR006171; Toprim_dom.
DR InterPro; IPR006154; Toprim_sub.
DR Pfam; PF02132; RecR; 1.
DR Pfam; PF01751; Toprim; 1.
DR SMART; SM00493; TOPRIM; 1.
DR TIGRFAMS; TIGR00615; RecR; 1.
DR PROSITE; PS01300; RecR; 1.
KW DNA repair; DNA recombination; Zinc-finger; Complete proteome.
FT ZN FING 57 72
SQ SEQUENCE 198 AA; 21689 MW; FC6F0E98E3933752 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 198;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
| | | | |
Db 101 HGLYHVLH 108

RESULT 11
SODC_HAEDU STANDARD; PRT; 199 AA.
ID SODC_HAEDU
AC Q59452; Q59449; Q59453;
DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
GN SODC.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000;
RX MEDLINE=97288949; PubMed=9143881;
RA Langford P.R., Kroll J.S.;
RT "Distribution, cloning, characterisation and mutagenesis of sodC, the
RT gene encoding copper/zinc superoxide dismutase, a potential
RT determinant of virulence, in Haemophilus ducreyi.";
RL FEMS Immunol. Med. Microbiol. 17:235-242(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=35000;
RX MEDLINE=97149276; PubMed=8996084;
RA Stevens M.K., Hassett D.J., Radolf J.D., Hansen E.J.;
RT "Cloning and sequencing of the gene encoding the Cu,Zn-superoxide
RT dismutase of Haemophilus ducreyi.";
RL Gene 183:35-40(1996).
RN [3]
RP SEQUENCE OF 100-186 FROM N.A.
RC STRAIN=35000;
RX MEDLINE=96118708; PubMed=7496539;
RA Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;
RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
RT from the eukaryotic enzyme, and not so rare after all!";
RL Microbiology 141:2271-2279(1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems. May play a role
CC in the interactive biology of organisms with their hosts and so
CC contribute to their capacity to cause disease.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X98737; CAA67289.1; -.
DR EMBL; U47664; AAB41293.1; -.
DR EMBL; X83125; CAA58206.1; -.
DR PIR; JC5718; JC5718.
DR HSSP; P24702; 2APS.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; sodcu; 1.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00087; SOD_CU_ZN_1; 1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW Antioxidant; Oxidoreductase; Metal-binding; Copper; Zinc; Periplasmic;
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 199 SUPEROXIDE DISMUTASE [CU-ZN].
FT METAL 92 92 COPPER (BY SIMILARITY).
FT METAL 94 94 COPPER (BY SIMILARITY).
FT METAL 117 117 COPPER AND ZINC (BY SIMILARITY).
FT METAL 126 126 ZINC (BY SIMILARITY).
FT METAL 135 135 ZINC (BY SIMILARITY).
FT METAL 138 138 ZINC (BY SIMILARITY).
FT METAL 173 173 COPPER (BY SIMILARITY).

FT DISULFID 99 195 BY SIMILARITY.
SQ SEQUENCE 199 AA; 21402 MW; 841D3210AB2BC06C CRC64;
Query Match 73.9%; Score 34; DB 1; Length 199;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 82 HDLAHGLH 89
RESULT 12
EX7L VIBCH STANDARD; PRT; 446 AA.
ID EX7L VIBCH
AC Q9KTW4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR VC0766.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E1 Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -!- FUNCTION: BIDIRECTIONALLY DEGRADES SINGLE-STRANDED DNA INTO LARGE
CC ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
CC INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE004162; AAP93931.1; -.
DR PIR; B82282; B82282.
DR TIGR; VC0766; -.
DR HAMAP; MF_00378; -; 1.
DR InterPro; IPR003753; Exonuc_VII_L.
DR InterPro; IPR004365; trNA anti_-
DR Pfam; PF02601; Exonuc_VII_L; 1.
DR Pfam; PF01336; trNA anti_-1.
DR TIGRFAMs; TIGR00237; xsea; 1.
KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 446 AA; 50542 MW; AA17369636A4BC9 CRC64;
Query Match 73.9%; Score 34; DB 1; Length 446;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVL 7 73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 155 HDILHVL 160

RESULT 13
EX7L_SALTY STANDARD; PRT; 449 AA.
AC Q8Z4Q1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR STY2753 OR T0345.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Hogue A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5' to 3'-
CC or 3' to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AL627275; CAD0214.1; -.
DR EMBL; AE016835; AAO68065.1; -.
DR HAMAP; MF_00378; -; 1.
DR InterPro; IPR003753; Exonuc_VII_L.
DR InterPro; IPR004365; tRNA_anti.
DR Pfam; PF02601; Exonuc_VII_L; 1.
DR Pfam; PF01336; tRNA_anti; 1.
DR TIGRFAMs; TIGR00237; xseA; 1.
KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 449 AA; 50720 MW; 511957DEC878F5D2 CRC64;

QY 1 HDLIHVL 7 73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 154 HDILHVL 160

RESULT 14
EX7L_SALTY STANDARD; PRT; 449 AA.
AC Q8ZN58;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR STM2512.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5' to 3'-
CC or 3' to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE008813; AAL21406.1; -.
DR StyGene; SG????; xseA.
DR HAMAP; MF_00378; -; 1.
DR InterPro; IPR003753; Exonuc_VII_L.
DR InterPro; IPR004365; tRNA_anti.
DR Pfam; PF02601; Exonuc_VII_L; 1.
DR Pfam; PF01336; tRNA_anti; 1.
DR TIGRFAMs; TIGR00237; xseA; 1.
KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 449 AA; 50613 MW; 85356CE8560E161E CRC64;

Query Match 73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVL 7
Db 154 HDILHVL 160

```
RESULT 15
EX7L_ECO57
ID EX7L_ECO57 STANDARD; PRT; 456 AA.
AC Q8XAB0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR Z3773 OR ECS3371.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005480; AAG57620.1; -.
DR EMBL; AP002561; BAB36794.1; -.
DR PIR; C91050; C91050.
DR HAMAP; MF 00378; -.
DR InterPro; IPR003753; Exonuc VII_L.
DR InterPro; IPR004365; tRNA_anti-.
DR Pfam; PF02601; Exonuc VII_L; 1.
DR Pfam; PF01336; tRNA_anti-; 1.
DR TIGRFAMs; TIGR00237; xsea; 1.
KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 456 AA; 51734 MW; 174EAE7F72EB3C37 CRC64;
-----
Query Match 73.9%; Score 34; DB 1; Length 456;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVL 7
||:|
```

Db 154 HDLIHVL 160

Search completed: November 12, 2003, 09:35:40
Job time : 11 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:32:27 ; Search time 28 Seconds
(without alignments)
73.729 Million cell updates/sec

Title: US-10-064-903-2
Perfect score: 46
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	% Match	Query Length	DB ID	Description
1	42	91.3	451	2 Q9R631	Q9r631 clostridium
2	42	91.3	1268	2 Q45851	Q45851 clostridium
3	42	91.3	1291	2 Q9ZAJ8	Q9zaj8 clostridium
4	42	91.3	1291	2 Q93G71	Q93g71 clostridium
5	42	91.3	1291	2 Q933K0	Q933k0 clostridium
6	42	91.3	1291	2 Q08077	Q08077 clostridium
7	42	91.3	1291	2 Q08GR96	Q08gr96 clostridium
8	42	91.3	1310	2 Q93N27	Q93n27 clostridium
9	39	84.8	397	16 Q92ML7	Q92ml7 rhizobium m
10	38	82.6	925	16 Q9JS16	Q9js16 chlamydia p
11	38	82.6	925	16 Q9Z8P5	Q9z8p5 chlamydia p
12	38	82.6	1278	2 Q57236	Q57236 clostridium
13	38	82.6	1280	2 Q9ZAJ5	Q9zaj5 clostridium
14	37	80.4	349	5 Q8IJV0	Q8ijv0 plasmodium
15	37	80.4	707	3 Q8X008	Q8x008 neurospora
16	36	78.3	105	10 Q8VXL7	Q8vxl7 fagus sylv

17	36	78.3	232	16	Q8X829	Q8x829 escherichia
18	36	78.3	237	13	Q8AWC9	Q8awc9 cyprinus ca
19	36	78.3	241	10	Q8VXL6	Q8vxl6 fagus sylv
20	36	78.3	489	9	Q9AZH2	Q9azh2 bacterioph
21	36	78.3	489	16	Q9CEA2	Q9cea2 lactococcus
22	36	78.3	773	10	Q9SHP2	Q9shp2 arabidopsis
23	36	78.3	1251	2	Q9K395	Q9k395 clostridium
24	36	78.3	1252	2	Q8KZM3	Q8kzm3 clostridium
25	36	78.3	1255	2	Q9FAR6	Q9far6 clostridium
26	35	76.1	312	16	O26068	O26068 helicobacte
27	35	76.1	312	16	Q9ZJ59	Q9zj59 helicobacte
28	35	76.1	426	5	Q8MYP4	Q8myp4 caenorhabdi
29	35	76.1	431	5	Q9XTZ9	Q9xtz9 caenorhabdi
30	35	76.1	500	16	Q9RVQ8	Q9rvq8 deinococcus
31	35	76.1	679	16	Q97SL8	Q97sl8 streptococc
32	35	76.1	737	16	Q8DR83	Q8dr83 streptococc
33	34	73.9	129	3	Q8TFT7	Q8tft7 ustilago vi
34	34	73.9	198	2	Q9RCP8	Q9rcp8 streptococc
35	34	73.9	198	2	Q9RCQ5	Q9rcq5 streptococc
36	34	73.9	198	2	Q9RCR0	Q9rcr0 streptococc
37	34	73.9	198	2	Q9R2M1	Q9r2m1 streptococc
38	34	73.9	198	2	Q9RCQ2	Q9rcq2 streptococc
39	34	73.9	199	16	Q8DV99	Q8dv99 streptococc
40	34	73.9	222	16	Q8EE87	Q8ee87 shewanella
41	34	73.9	225	5	Q8ILV8	Q8ilv8 plasmodium
42	34	73.9	261	5	Q8IFQ1	Q8ifq1 plasmodium
43	34	73.9	299	17	Q96YN4	Q96yn4 sulfolobus
44	34	73.9	385	10	Q8H3N0	Q8h3n0 oryza sativ
45	34	73.9	458	16	Q8FF64	Q8ff64 escherichia

ALIGNMENTS

RESULT 1

Q9R631 ID Q9R631 PRELIMINARY; PRT; 451 AA.
AC Q9R631;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type B light chain, BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92340509; PubMed=1634516;
RA Kurazono H., Mochida S., Binz T., Eisel U., Quanz M., Grebenstein C.,
RA Wernars K., Poulain B., Tauc L., Niemann H.,
RT "Minimal essential domains specifying toxicity of the light chains of
tetanus toxin and botulinum neurotoxin type A.";
RL J. Biol. Chem. 267:14721-14729 (1992).
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 451 AA; 51943 MW; 6C79FD488653EA71 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 451;
Best Local Similarity 87.5%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
Db 230 HELIHLVH 237

RESULT 2

```
Q45851
ID Q45851 PRELIMINARY; PRT; 1268 AA.
AC Q45851;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type F.
GN BONT /F.
OS Clostridium baratii.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93252228; PubMed=8486245;
RA Thompson D.B., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA Richardson P.T.;
RT "Nucleotide sequence of the gene coding for Clostridium baratii type F
neurotoxin: Comparison with other clostridial neurotoxins.";
RL FEMS Microbiol. Lett. 108:175-182(1993).
DR EMBL; X68262; CAA48329.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1268;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 219 HELIHLVH 226
|:|||||

RESULT 3
Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
ID Q9ZAJ8;
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bont protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13630; CAA73968.1; -.
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150840 MW; E4D3B0E46AB2E735 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 219 HELIHLVH 226
|:|||||

Q45851
ID Q45851 PRELIMINARY; PRT; 1268 AA.
AC Q45851;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type F.
GN BONT /F.
OS Clostridium baratii.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93252228; PubMed=8486245;
RA Thompson D.B., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA Richardson P.T.;
RT "Nucleotide sequence of the gene coding for Clostridium baratii type F
neurotoxin: Comparison with other clostridial neurotoxins.";
RL FEMS Microbiol. Lett. 108:175-182(1993).
DR EMBL; X68262; CAA48329.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1268;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 219 HELIHLVH 226
|:|||||

RESULT 3
Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
ID Q9ZAJ8;
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bont protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13630; CAA73968.1; -.
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150840 MW; E4D3B0E46AB2E735 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 219 HELIHLVH 226
|:|||||

Q45851
ID Q45851 PRELIMINARY; PRT; 1291 AA.
AC Q45851;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1436;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF295926; AAK97132.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 230 HELIHLVH 237
|:|||||

RESULT 5
Q933K0 PRELIMINARY; PRT; 1291 AA.
ID Q933K0;
AC Q933K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=593, and 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin.
KW Neurotoxin.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FA5A151 CRC64;
```


Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 6
Q08077
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE BONT/B.
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 7
Q8GR96
ID Q8GR96 PRELIMINARY; PRT; 1291 AA.
AC Q8GR96;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin.
GN BONTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RA Ihara H., Kohda T., Morimoto F., Tsukamoto K., Karasawa T.,
RA Nakamura S., Mukamoto M., Kozaki S.;
RT "Clostridium botulinum type B neurotoxin associated with infant
RT botulism.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB084152; BAC22064.1; -.
SQ SEQUENCE 1291 AA; 150574 MW; 0227CAEF4F58504D CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 230 HELIHLVH 237

RESULT 8
Q93N27
ID Q93N27 PRELIMINARY; PRT; 1310 AA.
AC Q93N27;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tetanus toxin (Fragment).
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RA Shumin Z., Dianliang L.;
RT "Cloning and sequence analysis of tetanus toxin gene.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF389424; AAK72964.2; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR006025; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1
FT NON_TER 1310
SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1310;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|||||
Db 234 HELIHLVH 241

RESULT 9
Q92ML7
ID Q92ML7 PRELIMINARY; PRT; 397 AA.
AC Q92ML7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative deaminase OR deamidase protein.
GN R02596 OR SMC02420.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Guzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";

```

RL  Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR  EMBL; AL591791; CAC47175.1; -.
DR  InterPro; IPR006680; Amidohydro_1.
DR  InterPro; IPR001130; TatD_DNase.
DR  Pfam; PF01979; Amidohydro_1; 1.
DR  PROSITE; PS01137; TATD_1; 1.
KW  Complete proteome.
SQ  SEQUENCE 397 AA; 43054 MW; B7DSF69C499CBE02 CRC64;

  Query Match      84.8%; Score 39; DB 16; Length 397;
  Best Local Similarity 87.5%; Pred. No. 20;
  Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
Db      326 HDLEHVLH 333

RESULT 10
Q9JS16
ID  Q9JS16 PRELIMINARY; PRT; 925 AA.
AC  Q9JS16;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT  01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE  CT234 hypothetical protein.
GN  CPJ0293 OR CP0465.
OS  Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC  Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX  NCBI_TaxID=83558;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AR39;
RX  MEDLINE=20150255; PubMed=10684935;
RA  Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA  White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA  Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA  Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA  Eisen J., Fraser C.M.;
RT  "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT  pneumoniae AR39.";
RL  Nucleic Acids Res. 28:1397-1406(2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=J138;
RX  MEDLINE=20330349; PubMed=10871362;
RA  Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA  Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT  "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT  from Japan and CWL029 from USA.";
RL  Nucleic Acids Res. 28:2311-2314(2000).
DR  EMBL; AE002208; AAF38302.1; -.
DR  EMBL; AF002546; BAA98503.1; -.
DR  TIGR; CF0465; -.
SQ  SEQUENCE 925 AA; 105601 MW; 61E8941E7C8FD620 CRC64;

  Query Match      82.6%; Score 38; DB 16; Length 925;
  Best Local Similarity 62.5%; Pred. No. 73;
  Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
Db      538 HDLLHITH 545

RESULT 11
Q9Z8P5
ID  Q9Z8P5 PRELIMINARY; PRT; 925 AA.
AC  Q9Z8P5;
DT  01-MAY-1999 (TrEMBLrel. 10, Created)
DT  01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  CT234 hypothetical protein.

```

```

GN  CPN0293.
OS  Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC  Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX  NCBI_TaxID=83558;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CWL029;
RX  MEDLINE=99206606; PubMed=10192388;
RA  Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA  Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT  "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL  Nat. Genet. 21:385-389(1999).
DR  EMBL; AE001614; AAD18442.1; -.
KW  Complete proteome.
SQ  SEQUENCE 925 AA; 105615 MW; 98E6098E7C8FD37D CRC64;

  Query Match      82.6%; Score 38; DB 16; Length 925;
  Best Local Similarity 62.5%; Pred. No. 73;
  Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
Db      538 HDLLHITH 545

RESULT 12
Q57236
ID  Q57236 PRELIMINARY; PRT; 1278 AA.
AC  Q57236; Q45863;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  Botulinum neurotoxin type F (BONT/F protein).
GN  BONT/F.
OS  Clostridium botulinum.
OC  Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC  Clostridium.
OX  NCBI_TaxID=1491;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=NCTC 10281;
RA  Hutson R.A., Collins M.D.;
RL  Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL  Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE OF 635-1000 FROM N.A.
RC  STRAIN=NCTC 1028;
RX  MEDLINE=94013372; PubMed=8408542;
RA  Campbell K., East A.K., Collins M.D.;
RT  "Gene probes for identification of the botulin neurotoxin gene and
RT  specific identification of neurotoxin types B, E, and F.";
RL  J. Clin. Microbiol. 31:2255-2262(1993).
RN  [4]
RP  SEQUENCE OF 1-27 FROM N.A.
RC  STRAIN=LANGELAND;
RX  MEDLINE=98404102; PubMed=9732534;
RA  East A.K., Bhandari M., Hielm S., Collins M.D.;
RT  "Analysis of the botulinum neurotoxin type F gene clusters in
RT  proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT  barati.";
RL  Curr. Microbiol. 37:262-268(1998).
DR  EMBL; X81714; CAA57358.1; -.
DR  EMBL; L35496; AAA23210.1; -.
DR  EMBL; X70821; CAA50152.1; -.
DR  EMBL; X99064; CAA67512.1; -.
DR  HSSP; P10845; 3BTA.
DR  MEROPS; M27.002; -.
DR  InterPro; IPR000395; Bontoxilysin.
DR  InterPro; IPR006025; Zn_MTpeptase.
DR  Pfam; PF01742; Peptidase_M27; 1.

```

DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin.
SQ SEQUENCE 1278 AA; 147073 MW; A1BE1318431D6918 CRC64;

Query Match 82.6%; Score 38; DB 2; Length 1278;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 227 HELIHALH 234

RESULT 13
Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.
AC Q9ZAJ5;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE BONT protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13631; CAA73972.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTPeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1280 AA; 147487 MW; D0F748976EBC222C CRC64;

Query Match 82.6%; Score 38; DB 2; Length 1280;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 227 HELIHALH 234

RESULT 14
Q8IJV0 PRELIMINARY; PRT; 349 AA.
AC Q8IJV0;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF10_0092.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014830; AAN35290.1; -.
KW Hypothetical protein.
SQ SEQUENCE 349 AA; 40775 MW; 9D7C20FEACED4464 CRC64;

Query Match 80.4%; Score 37; DB 5; Length 349;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 96 HELVHIVH 103

RESULT 15
Q8X008 PRELIMINARY; PRT; 707 AA.
AC Q8X008;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Related to hydroxyproline-rich glycoprotein.
GN B23H20.050.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL669988; CAD21077.1; -.
DR InterPro; IPR002965; P-rich extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 707 AA; 77817 MW; C49BC3C1A18D83F5 CRC64;

Query Match 80.4%; Score 37; DB 3; Length 707;
Best Local Similarity 62.5%; Pred. No. 87;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
|:|:|:|
Db 158 HDLLYLH 165

Search completed: November 12, 2003, 09:36:51
Job time : 31 secs